Goldsmiths
UNIVERSITY OF LONDON

e-scape portfolio assessment
phase 3 report

March 2009

This report has been commissioned by the British Educational Communications and Technology Agency (Becta)

ISBN 978-1-904158-99-8 (from Goldsmiths, University of London)

First published in Great Britain 2009 by Goldsmiths, University of London, New Cross, London SE14 6NW.

Additional copies of this publication are available from TERU, Design Department, Goldsmiths, University of London, New Cross, London SE14 6NW, price £30. Cheques, made payable to Goldsmiths College, should be sent with the order.

e-scape portfolio assessment

a research & development project for the Department of Children & Family Services (DCFS) and the British Education Communications Technology Agency (Becta)

phase 3 report

by the Technology Education Research Unit
Goldsmiths College University of London

Richard Kimbell
Tony Wheeler
Kay Stables
Tristram Shepard
Fred Martin
Dan Davies
Alistair Pollitt
Gillian Whitehouse

March 2009

Acknowledgements

TERU / Goldsmiths	Richard Kimbell, Tony Wheeler, Kay Stables, Chloe Nast
Science team	Dan Davies, Gareth Price
Geography team	Fred Martin, David Lambert, Tom Briebach
Schools and teachers	
design & technology	Saltash.net Community School, Dave Hayles Nancledra Primary School, Pauline Hannigan Clevedon Community School, Dave White Gordano School, Paul Wilson Priory School, Steve Yates Droitwich Spa High School, Andrew Millicheap Walsall Academy, Manjinder Sangha Burton Borough School, Helen Withington Abraham Darby School, Alan Beech Bedlington Community High School, Fiona Mather Coquet High School, Steve Tomlinson Duchess's High School, Craig Watson Jo Richardson Community School, Ifan Morgan Charles Darwin School, Richard Baker Uplands Community College, Roger Denholm Alexandra Park School, Ross McGill Ashfield School, Jo Hayes
science	Ralph Allen School, Frank Martinelli St Gregory's Catholic College, Jo Tidball Birley Community College, Ceri Howell
geography	Dyffryn School, Heather Anthony Y Pant Comprehensive, Dirk Sykes Llantwit Major School, Jane Jones
Additional trial schools	
	Seven Kings, London John Bentley School, Wiltshire Sir Bernard Lovell School, Bristol Hugh Sexey School, Somerset Hitchin Grammar School for Girls, Hertfordshire Stanchester School, Somerset Pencoed School, Bridgend Lavington School, Wiltshire Bath Spa University, Bath King Edward VII School, Leicestershire
Regional support team	John Chidgey, Paul Clewes, David Perry, Tony Lawler
Activity Administrators	Kay Stables, Ruth Wright, Tristram Shepard, Tony Wheeler, Richard Kimbell, Chloe Nast, Tony Lawler
Assessors / judges	Kay Stables, Ruth Wright, Tristram Shepard, Tony Wheeler, Richard Kimbell, Osnat Dagnan, Jenny Bain, Dave Hayles, Pauline Hannigan, Richard Baker, Paul Wilson, Manjinder Sangha, Alan Beech, Gurpal Thiara, Helen Withington, Andy Millicheap, Roger Denholm, Dave White, John Chidgey, David Perry, Niall Seery, Donal Canty, Tony Lawler, Paul Clewes, Ross McGill, Gillian Whitehouse, Frank Martinelli, Jo Tidball, Nicky Foster, Naresh Claire, Dave Harries, Dan Davies, Fred Martin, David Lambert, David Rayner, Alan Parkinson, Stuart Cousins, Justin Wolliscroft, Ruth Totterdell, John Lyon, Dirk Sykes

Becta	Niel McLean
TAG Learning	Karim Derrick, Declan Lynch
Handheld Learning	Graham Brown-Martin
Awarding Bodies	Dale Hinch (Edexcel), Patrick Craven & John Hill (OCR)

e-scape phase 3 Final Report

Contents

page

Executive Summary

overview

The e-scape project has progressed through phases one and two and this is the report at the conclusion of phase 3. The e-scape concept is built on a series of innovations in the world of performance portfolios and their assessment. Phase 1 (2004-5) was a 'proof the concept' in which we explored a range of hand-held tools and systems to establish that it was possible to link them directly to web-portfolios to capture live 'performance'. In phase 2 (2005-7) we built a prototype system for portfolio assessment in design & technology based on learners' performance in a 6 hour design task. The prototype worked very effectively and in phase 3 (2007-9) we were presented with two additional challenges:

- to establish the **transferability** of the system for assessments in a range of subjects other than design & technology
- to establish the **scalability** of the system and its applicability to coping with and managing national assessments.

There are two principal innovations in the e-scape system.

First, we have created a system in which school learners use hand-held digital tools in the classroom to create real-time web-portfolios. The hand-held tools are linked dynamically to their teachers' laptop – operating as a local server. This sends a series of tasks to the learners and 'hoovers-up' anything that they produce in response to them. The local server is enabled to upload – dynamically (in real time) - all the data from a class/group into a website where learners' web-portfolios emerge. In the phase 3 trials learners from 19 schools have created 350 design & technology, 60 science and 60 geography portfolios.

Second, we have created a web-based assessment system founded on a 'Thurstone pairs' model of comparative assessment. The web-based portfolios can readily be distributed anywhere at anytime – enabling multiple judges to scrutinise the portfolios simultaneously. The judging in phase 3 involved 28 design & technology judges, 6 science and 6 geography judges. All the judging was completed on-line in a short time-window and with extraordinarily high reliability (0.95). All these data can be downloaded into existing Awarding Body back-end systems for awarding and certification purposes.

e-scape elements

authoring tool

To enable the e-scape system to be as flexible as possible, we have developed an activity-authoring tool to enable teachers, examination bodies, or researchers to design their own activities. The tool operates on-line. Essentially the tool allows teachers to decide on any number of possibilities:

- the content area of the task (e.g. science/geography/design/drama)
- the overall duration of the activity
- the time sequence (e.g. 6 one-hour sessions, or a single three hour block)

• the sequence of sub-tasks that build up the overall portfolio
• the response-mode of learners (e.g. drawing/writing/photo/audio/video)
• the degree of flexibility in the timing of sub-tasks (controlled>flexible)
• the resource materials to be embedded (e.g. texts / images)

The tool allows teachers to design the activity – and modify it in the light of second thoughts or trial runs. It enables different sub-tasks to be selected for different learners – allowing teachers to personalise the activities to particular learners and their needs. Equally however for assessment activities, examination bodies can ensure that exactly the same activity is presented to all learners in all test schools.

the exam management system (EMS)

Once the activity has been designed, it can be transferred into the EMS, and it is from here that teachers manage the activity in the classroom. The EMS runs in the teacher's/administrator's laptop which operates through a local area network with wi-fi connectivity to learners' hand-held devices in the classroom / studio / workshop / laboratory.

At the start of the activity, the teacher activates sub-task 1 and this is sent to learners' devices. They work on it for the designated period (writing, drawing, taking photos etc), at which point the teacher's EMS 'hoovers-up' all their work back into the EMS and sends the second sub-task to learners. The closed network guarantees good data transmission between the teacher's and learners' devices.

Throughout the activity, the EMS enables the teachers to check that all the learners' devices are connected and operating properly – and gives a simple visual check on the battery state of each device.

The teacher works through the task merely by clicking on the big green 'next' arrow that takes the activity forward to the next subtask. Alternatively the teacher can choose to by-pass some of the sub-task steps using a drop-down menu that lists all the sub-tasks that were designed into the activity. At the end of the activity, the teacher uploads all the data from the EMS into the portfolio management system.

In the June 2008 e-scape round, teachers in 19 schools operated these activities with classes of 21 learners. The activities were:

science	- a 3 hr (one morning) activity
design & technology	- a 6 hr (two mornings) activity
geography	- a 5 hr (whole-day) activity

portfolio display

From the activities in the 19 schools, 350 design & technology portfolios, 60 science and 60 geography portfolios were uploaded into the web-space. The portfolios reflect the sub-task structure of the activities, but the display of them can be customised to facilitate the processes that follow.

For the purposes of the design & technology phase 3 portfolios, we intended to use a pairs judging process with multiple judges. We optimised the display so that all 25 sub-task boxes would fit as a single screen (without scrolling) on 20-inch monitors. This portfolio-at-a-glance arrangement allowed judges to scan the whole piece of work. Additionally however, judges can click on any of the elements in the boxes (e.g. photos / drawings / sound files) that automatically jumped to full screen images or play as voice/video files.

pairs engine

The pairs engine manages the assessment process. The system is based on a theory initially developed by Thurstone (1927) concerning the reliability of judgements. This theory was developed by Pollitt (2004) for inter-board reliability studies for GCSE and other school-based examinations; checking the reliability of the assessments that had been made. In phase 3 we have developed the system further so that the pairs judgements are the front-line assessments of the portfolios (i.e. there is no other assessment). We have developed the pairs engine to run this as an automated process.

The 'pairs engine' presents a judge with pairs of portfolios and the judge has to scrutinise the work and make a balancing holistic judgement about which of the portfolios represents the greater capability. For the design & technology sample we had 350 portfolios and 28 judges, each of whom made 130 paired comparisons. The geography and science samples were smaller and had judging teams of 6.

The judgement process is based on criteria, but these are not scored directly – but rather are interpreted by the judge into a single holistic judgement. At the outset the engine assumes that all the portfolios are of equal quality, so judges might well be presented with work that is radically different in quality. These judgments are easy and quick. As the data begins to build however, the engine starts to estimate a rank order and thereby presents judges with portfolios that are closer in quality. These judgements are more difficult and require the judge to look deeper into the portfolios to identify discriminating features.

Finally a complete rank order emerges – and with very high inter-judge reliability. For each portfolio the engine generates a 'misfit' statistic – essentially reflecting the amount of disagreement between judges that it created. Moreover, for each judge the engine generates a misfit statistic – reflecting the consensuality of that judge with the rest of the judging team. If either misfit statistic goes above an acceptable level, remedial actions are triggered. The remarkably high reliability of the judgement process (0.95) is explained by the fact that each portfolio is compared to approximately 20 others and is seen by more than 20 judges. The same levels of reliability were achieved with the science and the geography judging teams looking at their portfolios.

The out-turn data from the pairs engine can be fed into Awarding Body back-end systems for any subsequent awarding processes.

e-scape subject teams

The *transferability* element of our brief for phase 3, required that we demonstrate the viability of the e-scape approach to performance assessment in a range of subject settings beyond design & technology. As a result of discussions with DCSF, QCA, Becta and Awarding Bodies it was agreed that we would work with teams in geography and science.

The geography team was led by David Lambert, Chief Executive of the Geographical Society, who is also professor of geography education at the Institute of Education in London. He led a team based in Bath Spa University and the Sheffield offices of the GS. The science team was led by Prof Dan Davies at Bath Spa University in collaboration with the Centre for Science Education research group at Sheffield Hallam university.

The teams created activities in the authoring tool and developed them through a series of school-based trials. In the summer term of 2008 they each ran formal tests with groups of 21 learners in three schools: the science schools clustered around Bath and north Somerset and the geography schools in S Wales around Porthcawl and Port Talbot.

e-scape trials

The main design & technology sample of 13 schools was used to test the other principal element of the research brief, i.e. *scalability*.

In e-scape phase 2 we had demonstrated that the e-scape system did work, but it was based on two non-scalable elements:

i) the activity was hard-wired into the system and it was therefore not possible to evolve it or to make is useable for others who might want a different activity.

ii) the activity was run in the schools by the central TERU research team.

We took everything to the schools; set it all up; ran the tests with the groups of learners; and then left with the data. The schools merely provided us with teachers and groups of students to work with.

These phase 2 limitations clearly made that model of operation unrealistic as a national assessment system. To demonstrate the *scalability* of the e-scape system, in phase 3 we had to show that teachers, Awarding Bodies, and other subject teams could develop their own activities, and that the teachers could then run these activities in their own schools – unaided by the central research team. Managing the creation of learners' web-portfolios therefore became the central responsibility of the teachers, and we had to build the systems to enable them to do this autonomously. Subsequently we had to demonstrate that teachers could make all the necessary judgements that would result in the reliable assessment of these portfolios.

The authoring tool, EMS, portfolio display, and pairs engine were therefore developed with these priorities in mind.

The main design & technology sample was clustered into four geographic regions: Northumberland and the North East, the West Midlands, North Somerset, and London and the South East. We worked with teachers in developing the activity and trained them to use the systems for running the tests. In June 2008 we observed them operating the tests and uploading the portfolios. We then trained the judging team (a number of teachers from the trial schools + 15 others) and they all undertook the judging in October 2008. By November this was complete, and they reported back to us on the whole process. See section 14 for one teachers' complete account of the experience in his school. We have modelled the training requirements for teachers – both for running the activity and for the assessment process - and have realistic measures for what would be involved to implement the system nationally.

The combination of data generated within the e-scape phase 3 project confirm that it is both **transferable** to performance assessment activities in other subjects, and that it is a **scalable** system capable of roll-out for national assessment.

e-scape next steps

The principal aim of the next evolution of e-scape is to take it to an Awarding Body pilot that will involve national awards at 16+. This is planned to begin in September 2009 in association with at least one Awarding Body and run for the two years of normal 16+ examination courses. Discussions are underway with all the appropriate bodies - including Ofqual – to launch this final phase of e-scape.

The Awarding Body pilot project will run for two years in a nominated subject, with the e-scape component being one element of the assessment for that subject/award. Provided that the pilot is concluded successfully, it is planned that the system then moves towards commercial exploitation by Awarding Bodies and others.

e-scape phase 3 Final Report (Mar 2009)

The e-scape project has run through three phases starting in Jan 2005, and these grew from a previous project 'Assessing Design Innovation' that ran from Jan 2003 to Dec 2004. In order to contextualise this report on e-scape phase 3 we must begin by summarising the earlier work from which it grew.

1. background to e-scape phase 3

summary of 'assessing design innovation'

In 2003, the Technology Education Research Unit (TERU) at Goldsmiths University of London was asked by the Qualifications and Curriculum Authority (QCA) to undertake research to examine the extent to which - and the ways in which - innovation and team-work might be more fully recognised and rewarded in assessment processes, particularly within the General Certificate of Secondary Education (GCSE). The project '*assessing design innovation*' was launched in January 2003 and concluded in December 2004.

The principal outcome of that project was a developed portfolio assessment system that sat somewhere between a formal examination and a piece of coursework. It was designed to operate in 6 hours - typically 2 mornings - and presented learners with a design task that was to be taken through to a prototype. The outcomes of learners' work during this project were most encouraging. It was possible to demonstrate that different levels of innovation were identifiable in the work and that the best work was highly innovative. Critically, the consensus of teachers and learners was that the portfolio system acted as a dynamic force to drive the activity forward with pace and purpose. The data from the trials of this system is fully reported in the project report (Kimbell et al 2004).

Alongside this development, it is important to note a number of parallel strands of influence that combined to create 'project e-scape'.

Assessment for learning was becoming a major concern for educators. It places the teacher (rather than any external body) at the heart of the assessment process and presents teachers with large amounts of personalised-learning information to manage. Within this emerging field, we saw much value in exploring the use of digital systems to support teachers and learners.

In this digital context***, e-learning*** is a term that has emerged to describe a wide range of digitally enhanced educational experiences; from a straightforward internet search or the completion of a simple screen-based multiple choice question, to full blown multimedia managed learning environments providing access to complete courses. The DfES ***e-learning strategy*** identified the provision of a centralised e-portfolio as an important priority for reform, second only to the provision of the infrastructure to make it work.

In the context of design & technology alone, ***Awarding Bodies*** are responsible for the assessment of approximately half a million students annually using portfolios in which learners develop a design solution to a task of their own choosing, simultaneously telling the story of their development process. Approximately 50% of learners' GCSE marks are allocated on the basis of the quality of their portfolio. The Awarding Bodies responsible for these assessments – particularly at GCSE – are increasingly seeking to exploit the power of digital technologies.

This combination of influences led us at TERU to develop a proposal to QCA/DfES for a digital approach to portfolio assessment. Learning activities in design & technology studios and workshops were increasingly influenced by digital technology, and we believed that the portfolio assessment system that we had developed in the DfES "Assessing Design Innovation" project provided a useful model to explore the possibilities of extending digital working in design & technology into **digital assessmen**t of learners' performance. Project e-scape (e-solutions for creative assessment in portfolio environments) was established as a result of discussions with DfES, QCA and Awarding Bodies.

summary of e-scape phase 1 'proof of concept'

Phase 1 of project e-scape ran from January to October 2005 and was designed as a 'proof of concept' to see whether the kinds of technologies that existed at that time could be harnessed towards the creation of effective e-assessment portfolios.

The brief for phase 1 of project e-scape

"QCA intends now to initiate the development of an innovative portfolio-based (or extended task) approach to assessing Design & Technology at GCSE. This will use digital technology extensively, both to capture the student's work and for grading purposes. The purpose of Phase I is to evaluate the feasibility of the approach...' (QCA Specification June 2004)

The *proof of concept* operated at four levels; technological, pedagogic, manageability and functionality. Each of these four 'proof of concept' deliverables was explored in schools through a series of small-scale trials. We explored the system from both ends. At the classroom activity end, pedagogic priorities and the need to evidence capability dominated our concerns. But at the assessment end we were required to explore the manageability and functionality of an e-portfolio for assessment purposes.

Design studios and workshops are typically not full of desk-top or lap-top computers – which are often down the corridor in a different room – often an IT suite. Since we were concerned to explore the ***dynamic creation*** of an e-portfolio (rather than a sanitised one created through 2nd hand re-telling of the real story) we chose to use peripheral digital tools (e.g. digital pens, cameras, PDAs) that were capable of being used in the workshop/studio setting.

Specifically, the activity we were seeking to enhance was the 6 hour 'light fantastic' activity developed for the *assessing design innovation* project. This activity was capable of subdivision into a series of component parts, and – for the purposes of exploration with digital peripherals – we divided the activity into a series of 'work-parcels' – some focussed on supporting learners' designing and some on supporting teachers' assessment. We undertook a series of school-based trials between January and May 2005, with learners from year 6 to year 12. The second area of work

concerned the *technical systems* that would need to be in place for the learners to be able to develop their solution to the task in a *web-space* - accessible to the learners themselves, and their teachers, and (ultimately) to awarding body assessors.

The outcome of this phase 1 proof of concept was a body of digital work from learners and evidence from the associated e-assessment trials of that work. DfES, QCA and the Awarding Bodies were persuaded of the concept, and we were invited to take the project to the next stage.

The data and outcomes from phase 1 of the project are fully reported in the project report (Kimbell et al 2005). The work has also been published separately. See (Kimbell 2005) and (Stables and Kimbell 2006).

summary of e-scape phase 2 'working prototype'

Phase 2 of project e-scape (November 2005 – January 2007) involved the creation of a working prototype system. The system was to facilitate the designing activity in the classroom – using peripheral digital tools – and to allow all learners' work to be tracked and logged in a website for subsequent assessment by Awarding Bodies. The system was to be tested through a national pilot (principally with yr 10 learners) in the summer term of 2006.

Developing the system

The prototype was designed in association with two technology partners Handheld Learning (HHL) and TAG Learning. HHL are specialists in the use of peripheral digital tools and specifically PDAs. Phase 1 of the project had shown that PDAs would be good tools to focus on, principally because of their multi-functionality – for capturing drawings, photographs, writing and speech. TAG Learning have a strong track record in web-based portfolio creation for schools. For the phase 2 prototype we brought these two partners together and invited them to devise a system that would allow the handheld technology to 'speak' directly to the website, enabling us to track – in real time – the evolution of learners' portfolios in design workshops in schools.

Whilst the technology partners were working on the technical system, in TERU we worked on the classroom activity and the protocols that would render it manageable as a design activity worked largely through digital means. Whilst we worked from the 6 hour activity structure of the paper-based version in *assessing design innovation*, we modified it substantially to capitalise on the potential that is made available through the PDA. The drawing and writing tools can replicate paper based drawing and writing, and the camera removed the need for a separate device that we had previously used. But the speech tool was entirely new. For the first time we could capture the authentic voice of the learner at points through the activity. And throughout, we retained the importance of real materials – as learners endeavoured to create their own prototype solutions to the design task.

By May 2006 we had a working e-scape portfolio system, and all the resources needed to make it operate as an assessment system in schools. We launched the national pilot in 14 schools across the country through June and July 2006 and as a result accumulated 250 e-portfolios of year 10 learners' performance in the website. The

system worked remarkably smoothly and we are grateful for all the support and enthusiasm from teachers and learners.

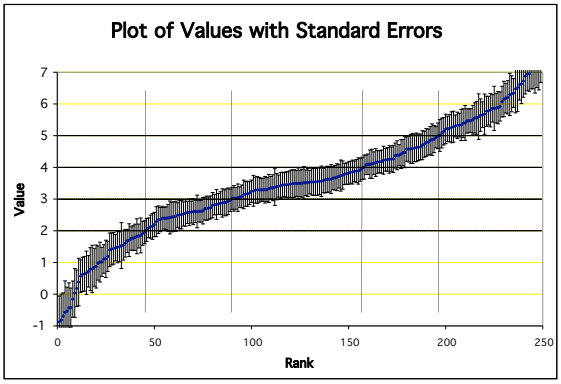
Making assessments

Assessing web-based portfolios can be done using the same systems as are conventionally used for assessing paper-based work, by allocating scores for elements within a rubric. However, for two related reasons, we have developed an alternative approach. First, because we are essentially concerned with assessing *performance-based capability*, an overview **holistic** judgement of the performance seemed more appropriate. Second, the fact that all the work sits in a website (and is therefore simultaneously available to as many assessors as we choose) opens the possibility of using 'comparative pairs' judgement (Thurstone 1927, Pollitt 2004) as a front-line methodology.

Essentially this involves teams of judges looking at pairs of portfolios and deciding (only) which – on the basis of agreed criteria - is the better / stronger piece of work. Then looking at another pair, and another pair, and another pair. Many paired judgements are made, enabling each piece of work to be compared to many other pieces, and the process undertaken by a team of judges so that many different judges see each piece. The combined effect of all this is two-fold.

First, using Rasch analysis, the mass of wins and losses for individual portfolios can be transformed into a rank order of all of the portfolios. Those at the top are those that win every comparison and those at the bottom have lost every time. In the middle are those that win half the time and lose half the time.

Second, since in our case each portfolio was judged at least 17 times (sometimes significantly more) and by 7 judges, the judging process renders highly reliable results. The standard error attaching to the placement of individuals within the rank order is significantly lower than would be the case in conventional portfolio assessment.



The judging process (including training of judges and 3 rounds of judging) was undertaken in September and October 2006 and the resulting data used for analysis of learners' performance. Whilst the pilot was principally designed to test the system itself, it was necessary to test learners within the system and the resulting data has proved interesting concerning for example the relationships in performance between designing on paper and designing digitally; between girls and boys performance; and in relation to different 'general ability' groups. Overall however, we were concerned to gauge teachers' and learners' reaction to the activity system in schools and the judges reaction to the assessment system.

Conclusions

The successful conclusion of phase 2 of project e-scape raised many issues of importance for the future of e-learning and e-assessment. We summarised these in relation to the four categories of research question that launched project e-scape.

Concerning **technological** challenges;

The key point here is that the whole system is driven by a remote server dynamically sending and receiving data to and from hand-held digital tools, putting the teacher in control of the sequences of the task and automatically building an evidence trail in the web portfolio.

Concerning **pedagogic** challenges;

The key point here is that everything we did for the purposes of collecting evidence for assessment **also** helped to scaffold the progress of the activity and the performance of learners.

Concerning the **manageability** challenges;

The key point here is the infusion of technology into activity. Real-time activity in studios, workshops, playing fields, theatres, science labs and the like, is typically not aligned with digital power. This power typically sits in splendid isolation in the shimmering purity of IT suites. In e-scape we have shown how the technology can get down and dirty and unleash its digital power where it is really needed. And in the national pilot we demonstrated that it was manageable.

Finally, concerning the **functionality** of the assessment system;

The key point here is that performance assessment is notoriously difficult, and at both ends of the system. It is difficult to manage the performance itself in ways that assure equity to all learners and it is difficult to ensure reliability in the assessment. Within e-scape we have created a system that operates right across this terrain. Learners and teachers report that it worked well at the performance end, and the data shows that it produced reliable statistics at the assessment end. The prototype has done enough to demonstrate that it is a functional system for assessment.

The data and outcomes from phase 2 of the project are fully reported in the project report (Kimbell et al 2007). Descriptions and analyses of this work have also been published separately. See for example (Kimbell 2007 [i]), (Kimbell 2007 [ii]), (Kimbell 2008 [i]), (Kimbell 2008 [ii]), (Kimbell and Pollitt 2008), (Stables and Kimbell 2007), and (Stables 2008).

2. structure of e-scape phase 3

brief for phase 3

The two major innovations in e-scape were (i) to create a system in which hand-held digital tools link dynamically to a website to create real-time e-portfolios, and (ii) enabling 'comparative-pairs' judging for reliable assessment. These two innovations were at the centre of the phase 3 brief from DfES, QCA, Becta, and Awarding Bodies.

However, in phase 1 and 2 we created a research-based set of tools and instruments – and demonstrated that they work in the context of research and development. It is a different thing however to establish that the ideas and procedures involved in the system can be expanded to become a basis for national assessment. The rationale for phase 3 was therefore based on moving the research-based work in phase 1 and 2 towards the priorities of national implementation. As part of this move, phase 3 was undertaken in close collaboration with national Awarding Bodies. This step into phase 3 therefore involved two principal extensions of the work.

1. Concerning specifically the ***transferability*** of the e-scape system (established in d&t in phase 1 and 2) to other subjects in the curriculum. After a good deal of discussion with QCA, Becta and Awarding Bodies, it was agreed that we would focus additionally on science and geography.
2. Concerning the ***scalability*** of the system. For the system to be operable for national assessment purposes it is not sufficient to have a system that can only be run by the research team (as we had in phases 1 and 2). It would be necessary to prove that it could be operated by teachers working in their own classrooms, studios, laboratories and workshops. And then we would also have to show that having created the web-portfolios, those same teachers could undertake the necessary assessments of the work.

research questions

Arising from the brief, there are three principal clusters of research questions that drive the research design.

1. Concerning the creation of structured coursework assessment activities (SCAAs):
• can teachers make effective use of the activity authoring tool to create their own SCAAs?
• can SCAAs be created – and worked effectively – in science and geography?

2: Concerning the operation of SCAAs and the construction of web-portfolios:
• can SCAAs (with all the associated technologies) be operated autonomously by teachers in their schools without the support of the e-scape research team?
• what degree of training is required by teachers to accomplish this?

• how does learner and teacher familiarity with the technologies influence the operation?
• how do the two principal technologies (Red-Halo / MAPS) operate in parallel within the classroom/school … and remotely using 3G communication?

3. Concerning the *assessment* procedures using comparative pairs judging:
• can the differentiated pairs methodology work through a single, self-sustaining, auto-updating programme?
• what training is required to enable teachers to operate as judges?
• can the output from the pairs judging process be incorporated into Awarding Body back-end systems for awarding processes?

The answers to the above questions will enable us to answer the key question of whether e-scape methodology provides a sustainable national system of assessment at GCSE.

After consultation with all parties involved in the earlier phases of the e-scape project it was decided that the contract should be managed through **Becta** in association with DfES (becoming DCSF), QCA, and Awarding Bodies

partnerships, roles and responsibilities

Phase 3 of project e-scape required a partnership of expertise across the following principal groupings:

TERU research team

The project was run by the e-scape research team at TERU, Goldsmiths University of London. We were responsible for managing all aspects of the project:
• establishing the subject working groups in geography and science, and monitoring and supporting their progress;
• liaising with Awarding Bodies concerning their procedures and contribution to the project;
• negotiating with technology suppliers to define and agree the e-scape system;
• identifying, selecting and purchasing the hand-held hardware on which the system would run for phase 3;
• user trialling of the alpha and beta systems created by the technology partners;
• managing the national pilot in June/July 2008, including all the liaison with schools;
• preparing the pairs assessment procedures, operating them and analysing outcome data; and
• preparing reports at regular intervals for Becta.

Technology partners (TAG Learning and Handheld Learning) have been central to the conception of e-scape from the start. Together they created the e-scape system that allows learners to operate in classrooms; created learner e-portfolios in a secure web-space; linked with Awarding Body back-end data management systems; enabled pairs judging by judging teams; enabled AB awarding process and created API

documentation to enable Becta to manage and control national standards in relation to software systems from other suppliers. These partners produced reports as necessary for the TERU research team.

Awarding Body partners (Edexcel and OCR) provided a valuable interface to schools and examination centres. They were responsible for ensuring that the emerging e-scape system remained viable as a school-managed technology and that the output from the e-scape system was manageable with their back-end data management systems.

Subject partners are comprised of the following:

Science:	Prof Dan Davies – Bath Spa University
	Gareth Price – Centre for Science Education; Sheffield Hallam University
Geography:	Prof David Lambert – London, Institute of Education and Geographical Association
	Fred Martin – Bath Spa University

Using the e-scape model of activity developed initially in design & technology, the geography and science teams were responsible for generating SCAAs in geography and science; trialling them in schools in paper-based and software modes, conducting a pilot in 'blind' schools; training judges for and undertaking the pairs judging of learners' performance and producing reports for the TERU team as necessary.

Assessment partners were of three kinds, two of which existed (at the outset of phase 3) within the QCA. The *curriculum* division of QCA had specialists in design & technology, science and geography that advised the project teams as their work on assessment activities moved forward. Additionally however the QCA had a regulatory division (now part of Ofqual) that is responsible for ensuring that proposed assessments are in line with assessment regulations currently in force. The innovative nature of e-scape SCAAs made it imperative that the project stayed in close contact with this regulatory arm of QCA. Finally, the project remained closely in contact with Alistair Pollitt, who was responsible for bringing to us the idea of 'comparative pairs' assessment. Pollitt managed the judgment pairings – and all the related data analysis – for phase 2 of e-scape, and in phase 3 advised the software development process as we built a 'pairs engine' that automated the system and enabled us to replicate on a bigger scale what his system managed for phase 2.

School partners have operated at various levels of the project. There are a number of schools that have been closely involved in the developments of phase 1 and phase 2. They have well-developed d&t practice; they are open to innovations in the classroom; they are enthusiastic about the pedagogic transformations that follow from the evolution of hand-held devices in classrooms; and they have a supportive school management group enabling the teachers (for example) to arrange for us to have access to groups of learners for trialling purposes. It is in these 'development' schools that SCAAs (and the overall e-scape pedagogy) have evolved through an iterative process of trials and modifications. Both the teachers and the learners in these schools

know that they are co-researchers with us. At the opposite end of the spectrum, national pilot schools are broadly representative of schools nationally. They were initially selected by Awarding Body partners and thereafter visited by the team to ensure that there were no reasons (technical or other) to exclude them. Because of the limits imposed by the amount of kit and the number of researchers in the team, we attempted to cluster schools in geographic regions.

methodology and project plan

A research approach

Underpinning this research and development project is an approach to performance assessment that originated in our Assessment of Performance Unit (APU) research of the 1980s and was developed by Stables and Kimbell through a series of subsequent projects. We christened it the 'un-pickled portfolio' approach (see Stables & Kimbell, 2000) in the sense that performance was tracked in real-time without the normal benefit of hours and weeks of reflection time (pickling) in which to pretty-up the portfolio. In our un-pickled versions, learners' portfolios emerge as the raw, real-time 'trace-left-behind' by their activities. The transition from 'timeless' portfolios (in d&t examinations in England and Wales, frequently prepared over many months) to 'real-time' portfolios is significant in several ways, but centrally the issue concerns direct evidence of process-centred capability.

Real-time approaches for the collection of performance data have been used for some years in the worlds of management and computer science, where system evaluation (and indeed assessment) has entailed the forensic analysis (step-by-step / minute-by-minute) of sub-elements of the performance of systems (e.g., Calvez, Wyche, & Edmundson, 1993; Mohd Adnan, 1993). And in the learning context, using interactive computer systems, Hillary (2005) has developed a real-time

> … interactive tool for researchers to track and examine the activity patterns of users … (Hillary 2005)

It seemed to us – as early as the APU work of 1985 - that since we claim that *process* (i.e. designing) performance is central to what we do and value – we should develop exactly this forensic approach to the creation of design portfolios.

The first step in this process is to create an activity through which learners progress from a task statement towards a design prototype. The *choreography* of the activity is informed partly by pedagogic priorities (what will work well as an activity) and partly by assessment priorities (what trace will be left behind by it – and how can we enrich it). Sometimes these priorities work well together and sometimes they conflict … leading to a good deal of trialing and optimisation in the classroom/studio/workshop. The pill storing/dispensing activity from phase 2 evolved into a 6-hour activity that moved through 24 sub-tasks that varied in time allocation from 5 mins to an hour. The student

responses to these 24 sub-tasks gradually build in the portfolio to create the evidence trail that is left behind by the activity.

sequencing the work

Phase 3 operated broadly through five stages – reflecting the 5 academic terms of the duration of the project.

Autumn term 2007:

• Geography and science teams created paper-based SCAAs and trialled them in schools

• Central TERU D&T team created a DIY system for teachers so that they could create and run SCAAs autonomously in their schools

• Technology partners created the alpha version of the software system and authoring tool

Spring term 2008

• Geography, science and d&t SCAAs were loaded into e-scape system and trialled in schools

• Technology partners (using the trials experience) refined the software system to beta version and finally to 1.0 release version for the national pilot

• TERU team prepared training materials for teachers, conducted the training and arranged final plans for the national pilot

Summer term 2008

• The national pilot operated in schools. The design of the design & technology element of national pilot was aimed at informing the 'scalability' question and therefore had a number of variables. Central to these was the classification of schools into two categories:

"hot" schools had previous experience of the e-scape system and had the technology for an extended period in which to embed it into their curriculum

"cold" schools were selected only for national pilot purposes and were therefore running the e-scape task 'cold'

These different groups allow us to report on the teacher and learner training and familiarity issues relating to the *scalability* of the e-scape system.

• during this period the technology partners delivered the 'pairs engine'; the judging system software

• TERU recruited the judging team and prepared training materials for judging training

• Subject teams conducted local pilots of their activities to inform the question of transferability.

Autumn term 2008

• TERU undertook the training of the d&t judging team and the science and geography partners, who then recruited and trained their own judging teams

• The technology partners and Awarding Body (AB) partners ensured the effective communication between e-portfolio system and AB back-end systems

• all work (geography, science and d&t) was put through the judging process
• all learners' performances were analysed

Spring term 2009

Data from the pilot was analysed and the final research report was compiled during this final period of the project. In this report we describe the whole e-scape system and demonstrate how it can be used as a valid and reliable approach to coursework assessment in schools. In the process we show how the system operates effectively at the classroom level, enabling learners to create e-portfolios of their performance. Additionally we show how the e-scape system can be operated as a national system of assessment that will:

• enable Awarding Bodies to register candidates in their normal way
• select pairs of portfolios for comparison and allocate them to particular judges
* present the scripts in a suitable interface to facilitate paired judging
• be a responsive / dynamic tool such that any judge's responses immediately update the system and adjust the next sets of pairs to be selected
• operate with maximum efficiency so as to arrive at a reliable rank-order of candidates with the minimum judging interventions
• create a rank-order in ways that make it susceptible to analysis and awarding through Awarding Body systems, including identifying and validating grade boundaries
• facilitate the identification of non-consensual judges and problematic portfolios (where judges disagree)
• produce reports of system status at designated points through the cycle.

The overall aim of phase 3 was to model a scalable national system for secure assessment. Accordingly the national trial demonstrated a complete run through of the combined systems. We have completed this full simulation of the e-scape coursework assessment system in a way that enables Awarding Bodies to identify how they might take it forward into formal GCSE pilot programmes acceptable to QCA/Ofqual.

A schematic of the plan for the whole project is presented in the Gantt chart in Appendix 2. Additionally Appendix 1 shows the deliverables for this project that were agreed with Becta at the outset.

3. establishing the project

selecting schools

The schools that we worked with fall into one of two categories 'hot' and 'cold'. In reality these categories are not watertight and there is some blurring at the edges. The project plan required that we trial materials (and the software system) in schools through February and March '08 and that we then conduct the main survey in schools in June and July.

The trials of system were done principally in 'hot' schools – the definition of which is that

• they have well-developed practice in product design courses at GCSE
• they have run e-scape activities in the past (mostly in the 2006 phase 2 survey). and/or
• they have been development and trialling centres for our projects in the past

The rationale for having hot schools is that they understand what is involved in research with TERU and specifically what will be required of them in terms of designing, creating and running e-scape activities. They are then well placed to lead the development process with us. They are our most critical and creative partners – always wanting the system to do more than it currently can – and always helping us to improve on it. The demands on these schools are consequently very considerable and our collaboration typically depends on the good-will and availability of key personnel in the departments. This makes our partnership vulnerable to the vicissitudes of sickness and other unavoidable absence.

There were originally 6 hot schools in phase 3, five of which are mixed comprehensives, and one is a primary school. However due to staff sickness (in one school) and pregnancy (in another) we ended up with four schools.

- Alexandra Park School, London.
- Ashfield School, Nottinghamshire
- Dixons City Academy, Bradford (withdrew)
- Nancledra CP School, Cornwall
- Saltash Community School, Cornwall (withdrew)
- Uplands Community College, E Sussex

The reason for the primary school (in what is essentially a 16+ assessment project) was the same as it had been in phase 2, where we were able to show that primary teachers and learners are just as able to accommodate the practical and intellectual challenge of e-scape as were their secondary school counterparts. We were also keen – in phase 3 – to explore the cross-curricular potential (geog / science / d&t) that is so much easier to facilitate in primary schools.

The cold schools list was finalised somewhat later as they were not needed for the early trials. We explored lists supplied to us by Awarding Bodies, LEA advisory experts, teacher education connections, professional associations and our own networks of

teachers and schools. The number of schools overall was a product of the amount of kit we had to distribute. We acquired 10 sets of kit, so at any time we could run 10 schools simultaneously. These 10 sets were subdivided in the research design as follows:

- one set to science
- one set to geography
- four sets to the d&t hot schools
- four sets to the d&t cold schools

Potential cold schools were all visited during December '07 and January '08 by the TERU team members and the final list was confirmed by the end of February 2008. In our preliminary visits, we ensured that teachers realised the full extent of the work they were committed to – essentially running the e-scape activity for themselves with us as observers. This is radically different from the demands of phase 2, in which we (TERU) ran the activities with the teachers as observers. It is to this end that we developed the DIY escape kit (see below in sect 4)

The four sets of cold school kit would enable us to manage 12 cold schools and for efficiency reasons we decided that they should be clustered into geographical regions

- the north east – Northumberland / Newcastle
- the south west – North Somerset / Avon
- the west Midlands – Shropshire / Wolverhampton / Telford
- London and the South East

The principal reasons for clustering cold schools were as follows. First, the schools could operate as a local support group for each other. Second it helped us to identify a local support person who was more immediately available than we were in London. Third, the activity takes 3 days to run (1 day training the learners then two days of activity) and - with travelling - this involved nearly a week away for us. We found in phase 2 that it was more efficient for us to travel - as a team - to a region and base ourselves there for the week. In this way we could better support the research process for each other.

The regional clusters of cold schools emerged as follows:

West Midlands

- Droitwich Spa High School
- Walsall Academy
- Burton Borough School, Newport
- Abraham Darby School, Telford

Northumberland

- Bedlington Community High School
- Coquet High School (Technology College), Amble
- Duchess's High School, Alnwick

South West

- Priory School, Weston Super Mare
- Clevedon Community School
- Gordano School, Portishead

London & SE

- Charles Darwin School, Biggin Hill
- Jo Richardson School, Dagenham
- Seven Kings, Ilford

The total school sample for design and technology therefore evolved to be 4 hot schools and 13 cold schools. These 17 schools provided groups of 21 learners for the main e-scape d&t test activity, giving us a target sample of 357 learners

establishing the geography and science teams

We have explained above that the central rationale for activities in other subject areas is to explore the ***transferability*** of the e-scape approach from design & technology to geography and science. The two research teams were recruited and commissioned by TERU to undertake the work in association with us. So whilst there is a degree of difference in their work - arising through the different content, procedures and cultures of subjects, there is also a great similarity. This similarity arises from the brief that we agreed with the teams which is summarised in the following framework.
(see Appendix 3 for the full Memorandum of Agreement)

The task

The following broad guidelines will be used to steer the development of the assessment activities. They are based on the existing e-scape task framework

- activities should reflect the very best *active-learning* models that exist nationally within geography
- the *duration* of activities is determined by the need for *authenticity* (e-scape d&t = 6 hrs: two mornings)
- activities should be seen as *valid* activity by the geography community
- activities should encourage elements of *group/team contribution* in support of individual performance
- activities will be based around an *overriding task/purpose* within a geography context
- activities will be *thoroughly contextualised* for learners, providing meaning and purpose for the activity
- activities will unfold through a *structured and interconnected* series of *sub-tasks* of varying length
- activities will create *standardised demands* on learners but encourage *diverse responses* from them
- the encouragement of *individual creative responses* will need to be signalled and supported
- activities will be resourced by whatever *equipment & resources* are necessary for authentic activity
- learners' work will be undertaken on *hand-held digital tools* using the *e-scape interface*
- learners *e-portfolios* will emerge in the secure *e-scape web-site*, from where it will be assessed.
- whilst the principal target is *year 10 learners*, we are interested also to explore activities with *year 4/5/6*
- activities will be developed through small-scale trials in appropriate schools

The development process began with a formal induction / training process at a 2 day residential training session in Rugby (June 2007) and thereafter through a series of formal and informal meetings between the teams. In addition to these meetings we have attended as observers a series of trials of materials in schools.

early issues arising through geography

The geography experience gave rise to some potentially big issues that had not arisen through design & technology. The first of these concerned the debate (in the geography community) about ***fieldwork***. It is probably rightly thought that fieldwork is critical to good geographical activity, since it involves groups of learners getting out into the real environment to study it. The issue therefore for the geography team was whether to incorporate fieldwork into the e-scape activity. Was it to be based on real live data collected by learners or on second hand data from other sources?

The second issue concerns the centrality of ***place*** in geography. Whilst an activity about "re-branding Porthcawl" makes sense to children in Pencoed – which is just down the road and familiar to most of them – it might be very different for children in Manchester, Hull or London. Why would they be interested in re-branding a distant town in south Wales? They might though be more interested in re-branding a local town of their own. And this suggested that the e-scape geography tasks might need to be written in a form that allowed local customisation by the teacher. And this in turn would require different sets of resources supplied by the teachers. So whilst there would be a generic activity that would be the same across the country – the specific activity would appear different in each school/town.

early issues arising through science

The two tasks in science are both broadly set in the context of road safety … one focussed on the physics of mass / acceleration / impact damage / stopping distance etc, and the other on the biology of reaction timing and specifically on examining the effects of a number of factors that affect the speed of reaction of a driver.

Again a couple of critical issues emerged through the early work, and the first of these was **time**. The task development process was closely attuned to trials in schools, and specifically in the context of one-hour science lessons. So the activity – which was intended as a complete 3 hr task – was trialled initially in three sections. We were confident however that whilst schools might be inconvenienced by having to create whole-morning sessions to run the activity, they would be fully rewarded by the result.

A second issue concerned resources. The science team was anxious to ensure that schools did not have to provide extensive resources to run the activity. And whilst the value of, for example; 'handling collections' were well understood in design & technology, they was not exactly in the main stream of science lessons. Nevertheless, having explored these features of the d&t activity, the science developers adopted them and were enthusiastic about their effects.

4. building the software system

In phase 2 of the project we had developed a working prototype of the e-scape system in collaboration with the awarding bodies and our two technology partners TAG Learning and Handheld Learning. While functional as a prototype, the phase 2 system lacked some key components and in phase 3 we refined this into a more robust, manageable, scalable and transferable system.

authoring tool

In particular we needed to allow researchers, teachers and examiners to create, modify and share e-scape activities. In the phase 2 prototype, this was not possible without the assistance of a programmer. For phase 3, we specified an activity-authoring tool that became known as the SCAA creation tool (structured coursework assessment activity).

management system

In the phase 2 prototype it had been necessary for a member of the research team to run the e-scape activities. In phase 3 it was important to show that teachers could run the e-scape activities themselves in their own classrooms. To do this we needed a much more “user-friendly” management system with clear, unambiguous controls. This became known as the EMS (exam management system).

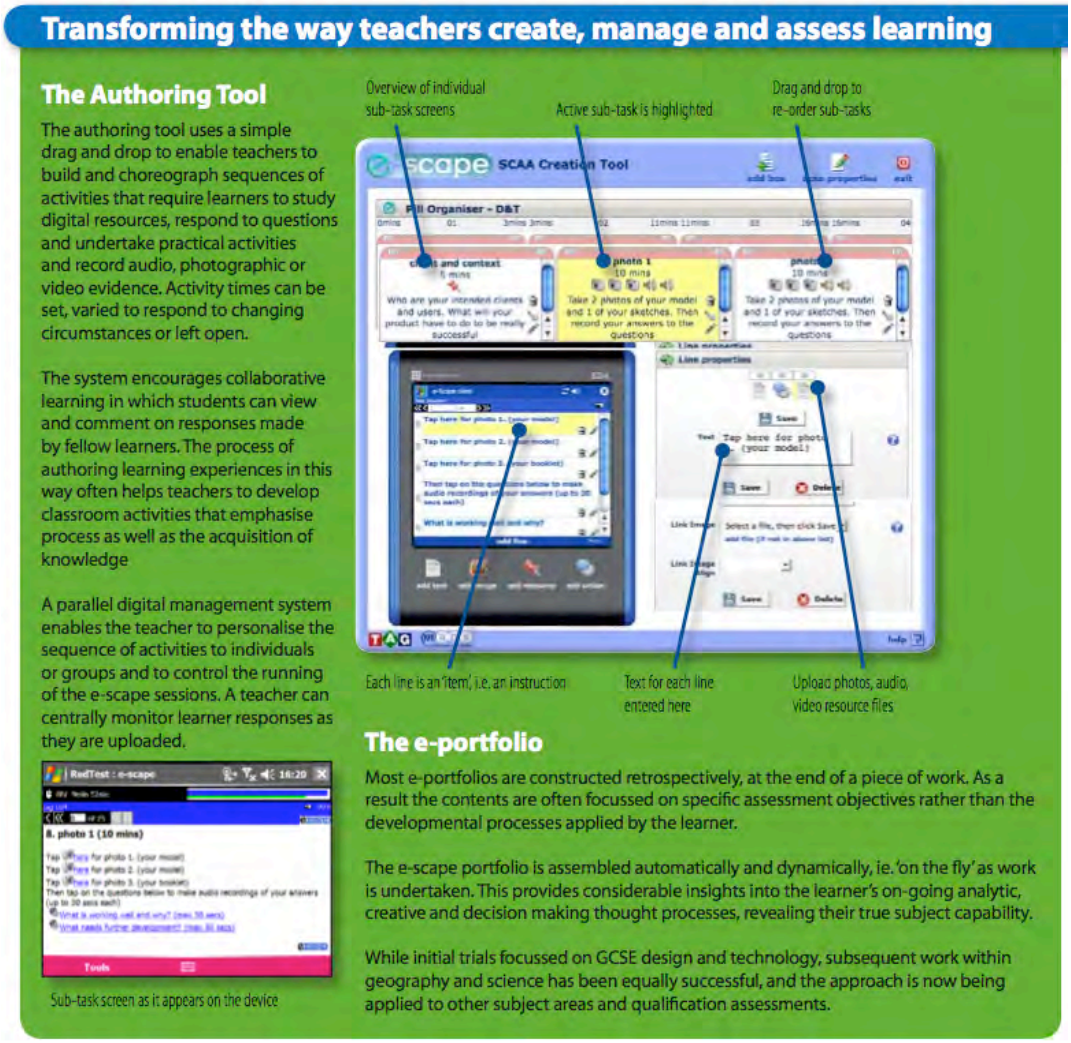
portfolio viewer

The experience of working with e-scape portfolios in phase 2 had suggested a number of refinements to the portfolio display system including the addition of video, spreadsheet and mind mapping data capture as well as the need to rotate, enlarge and sharpen some photographic images.

pairs engine

The phase 2 judging process presented the learners work on screen but the assessments were collected on paper and transferred manually to another system to be collated and analysed. To facilitate the pairs comparison assessment system across a wider community of judges we needed to develop a single integrated online assessment system. This became known as the “pairs engine”.

In order to explain the basics of how the e-scape system works, we prepared a leaflet introducing the components of the system. (For the full contents see Appendix 4.)



Section from the e-scape leaflet

the development process

The perennial problem with the development of innovative digital system has been that users do not understand what a particular technology can do until they have used it and, although technologists know about technology, they often misunderstand the needs of the users. Bridging this disconnection is difficult and requires a partnership that values:

- individuals and interactions over processes and tools
- working software over detailed documentation
- customer collaboration over contract negotiation
- responding to change over following a plan

(ref: http://en.wikipedia.org/wiki/Agile_software_development)

Agile development methodology

In common with the previous phases of the project we employed an "Agile" development process. In research situations where the outcomes of pedagogic interventions are unsure it is inappropriate to lock down the technology required to support the research at the outset. If detailed too early, it is often the case that the technology stifles and restrains the research. To ensure the effective management of a

more open approach to technology procurement in these circumstances, rather than specifying in detail the exact functionality of the technology, it has become widely acceptable to detail the process and user requirements of the technical development instead. In this way the technology can become a more integrated and supportive part of the research development process.

The agile approach anticipates a continuous stream of interactions with the clients, delivering components frequently for evaluation and testing. For the e-scape project we achieved this by breaking the system down into a number of connected but self-contained modules. The research team worked with TAG and Handheld Learning to specify the initial functionality for each of the component parts and this was recorded in the design specification. (design document - see Appendix 5.)

Importantly this was not then seen as a constraint but a starting point, and throughout the development changes were welcomed, agreed and incorporated into the specification that reflected the active development of the system as it was employed in the classroom with a widening range of users and circumstances.

trialling

Typically this *agile* implementation followed a similar pattern throughout involving three levels of trialling. Initially the technology team would engineer a component to a point where it functioned. This would then be trialled with the research team and changes would be agreed and implemented before a second level of trials, usually with students from the university taking the roles of pupils and teachers. This second level of trialling allowed the research team to make more detailed observations alongside the technologists. Further changes were implemented at this stage before taking the modules to a final level of trialling in individual schools where further modifications emerged.

Usability review

In addition to the observations of the technical and research teams we employed a usability expert, Penn Smith, to review the first full trial of the authoring tool, management system and pupil handset software at Seven Kings School in Ilford. Smith worked with the development teams before the trial and attended both sessions of the first full phase 3 trial. The exam management system (EMS) is critical to the effective running of the activity and was the main focus of attention. Smith's observations, together with those of the researcher who ran the session, and the classroom teacher, provided a framework for redesigning the management system interface. (For the full usability report see Appendix 6.)

managing the integration of two systems (MAPS/Redhalo)

Although TAG and Handheld Learning worked together on phase 2 of e-scape, the scope of the new system was sufficiently different to warrant a completely new approach. As such the phase 3 system was developed from a completely fresh code base – there was no overlap from the previous project.

The first phase of the development process involved a series of meetings between TAG and Handheld Learning which saw the generation of a detailed specification document which presented in detail the workings of the new system. Through these meetings a thorough and clear understanding of the new system was laid out.

The solution from Handheld Learning incorporated their full RedHalo system. RedHalo is a server – client system from Handheld Learning that runs on a range of popular mobile devices. The main feature of RedHalo is its ability to synchronise devices with a central server. In order for this to run it was necessary to install the RedHalo client on each device. Once the RedHalo client was installed it was then able to automatically update the software on the devices – including updates to the RedHalo client if they were required.

Communication between the MAPS/EMS system and the RedHalo server was managed through the development of a specific API (Application Programming Interface). The API was sufficiently detailed to allow the EMS to control all the management functions of running and updating sessions.

One of the benefits of using the API was that the physical location of the software was no longer an issue. This meant that the MAPS/EMS system could talk to the RedHalo server at any location. When running the system over the internet, the machines were quite separate. When running the system in the classroom with a local laptop server, the MAPS/EMS system and the RedHalo server ran on the same machine. Because all communication was handled through the API layer, the same code was able to run in each scenario and meant that both TAG and Handheld Learning were able to work on one system.

Due to the detailed nature of the specification drawn up at the start of the project and the in-built flexibility of both systems, there were no major issues relating directly to the API during the running of the system.

The one issue that did arise during the development process was the fact that the RedHalo accounts that ran on the server were permanently locked to the device on which the corresponding client was initially installed. One of the early plans for the project was that each student would be able to have their own account and log on using their personal details. This became logistically impractical and would have been difficult to implement with RedHalo given that the same devices were being moved between schools and would have had to have their account details updated to enable this.

Inter-operability APIs and other management systems

The integration of the pre-existing AB (Awarding Body) system into e-scape was quite straightforward. The AB system is similar to one used in the OCR and AQA awarding bodies in live qualification contexts. As all this development was undertaken in-house, TAG were able to utilise the structure of the existing Awarding Body system to enable it to handle comparative pairs assessment.

Within MAPS there already exists a mechanism (a combination of XML schema's and API's) to enable it to describe the contents of a student portfolio of work and submit that to an Awarding Body. The MAPS system has an extra layer of detail where it also

matches the submitted work with a detailed specification for that qualification and its associated mark scheme. For e-scape this was superfluous as there is no need for a detailed mark scheme in the context of holistic judgement.

In the final e-scape system, the MAPS users were able to draw down an Awarding Body approved task. This task was then assigned to students in the school. Work from the local laptop was imported in the form of a zip file. These imported portfolios were then submitted to the e-scape generic Awarding Body system. The mark scheme used for the mock e-scape qualification was set such that the Awarding Body system recognised that the marking approach was that of comparative pairs.

Once the work had been submitted to the system it was then possible for the Principal Moderators to assign judges and to undertake the comparative pairs assessment process.

handsets (PDA/netbook/phone)

There have been significant and ongoing technical developments in handheld technologies throughout the phases of the e-scape project. In particular there has been a continuing convergence of handheld gadgets such as music players, cameras, portable game consoles and phones into “smart” combination handsets, typically either smart phones, handheld gaming machines or multi function MP3 players.

At each stage of the project we have questioned pupils about their personal experience of handheld devices. In phase 1, while almost every pupil had a mobile phone, most were surprised that anyone would want a camera included in it. By phase 2, many pupils had better cameras in their phones than we could offer in the PDA we provided, but they had no idea what “smart” features were and few had experienced touch screens.

By phase 3 the ubiquity of the i-phone, i-pod touch and gaming devices such as the Nintendo DS had altered the perception and expectation of pupils and while they did not necessarily all own a touch screen device they knew what they were and had all used one.

In phase 2 our wide range of functional requirements for the handset (wi-fi, camera, touch screen, sketch tool, long battery life, etc) restricted our choice to a single device, the HP i-paq rx3175 PDA and we had to accept a degree of compromise in each of the features in order to have them all. (i.e. the camera did not have particularly high resolution, the screen size was a bit small, etc)

In phase 3, thanks to their industry contacts, Handheld Learning provided access to products which had not been released to the general market and it became clear that mobile telephone handsets, rather than PDAs, were likely to provide more of the features we required. Two devices in particular stood out. The HTC advantage (distributed by T-mobile in the UK as the Ameo) and the Toshiba Portege. We acquired samples of both and over the summer tried these out with groups of young people.

The Portege is smaller and looks more like a traditional mobile phone with a slide out keypad. The Ameo is larger, with a detachable screen and looks more like a mini

computer. Both are mobile phones running the Windows mobile operating system and have similar feature sets (camera, audio, touch screen etc). The pupils agreed that they would expect to use a device like the Ameo in school and would want the Toshiba for their personal use. In particular because of the bigger screen and the better camera we chose the Ameo for the majority of the trials, we also invested in a class set of the Toshiba phones to establish whether there were identifiable handset effects on pupil performance.

purchasing hardware

For the 1st 7 months of phase 3 (Feb – Aug 2007) we were running trials on a range of hand-held devices that we felt might be appropriate for the task in phase 3. Having resolved the selection down to the two devices, we decided to acquire both; the bulk of the devices (9 sets of 24) being the HCT Advantage, but with a small number (one set of 24) of the Toshiba.

T-mobile Ameo (re-badged HTC Advantage)
- 5" touch screen (800x480 pixels)
- detachable QWERTY keyboard
- 3 Megapixel Camera
- USB and VGA ports
- 3G (mobile broadband)
- Wi-Fi
- Bluetooth
- GPS receiver
- MS Windows Mobile 5

Tosheba Portege 700 smart phone
- 3" touch screen (800x480 pixels)
- slide out QWERTY keyboard
- 2 Megapixel Camera
- USB port
- 3G (mobile broadband)
- Wi-Fi
- Bluetooth
- MS Windows Mobile 6

We used the Government Procurement Service – www.ogcbuyingsolutions.gov.uk as a first line of enquiry. When we had identified the requirements we acquired estimates from suppliers and moved ahead to selection from there. Using this process, the

suppliers - T Mobile - were approved by Becta and along with the hardware they also supplied the connectivity contract for the 3G exploration.

In relation to the supplementary computer equipment for the classroom-based activities it proved difficult to specify our requirements within the OGC system - partly because of the combination of soft and hardware that it involved. So as a 2nd option, we used the London Universities Purchasing Consortium. From there we asked four suppliers to quote for the listed equipment. (The quotes and process are outlined in the chart in Appendix 7.).

connectivity (Wi-Fi/3G)

Given the wide variability of school networks in terms of overall efficiency and capacity, coverage across the school site and peak usage, we had determined that we would need a connectivity system for e-scape that was independent of the school systems over which we had no control.

The main difference between the hardware used in phase 3 was that the devices we chose were mobile phones rather than PDA handsets. This meant that we could explore the possibility of networking through the mobile phone system rather than setting up our own portable Wi-Fi cluster as we had in phase 2.

Although the standard voice and text phone network was not designed to carry significant packets of data, the emerging 3G networks were promoting the ability to carry video to individual handsets.

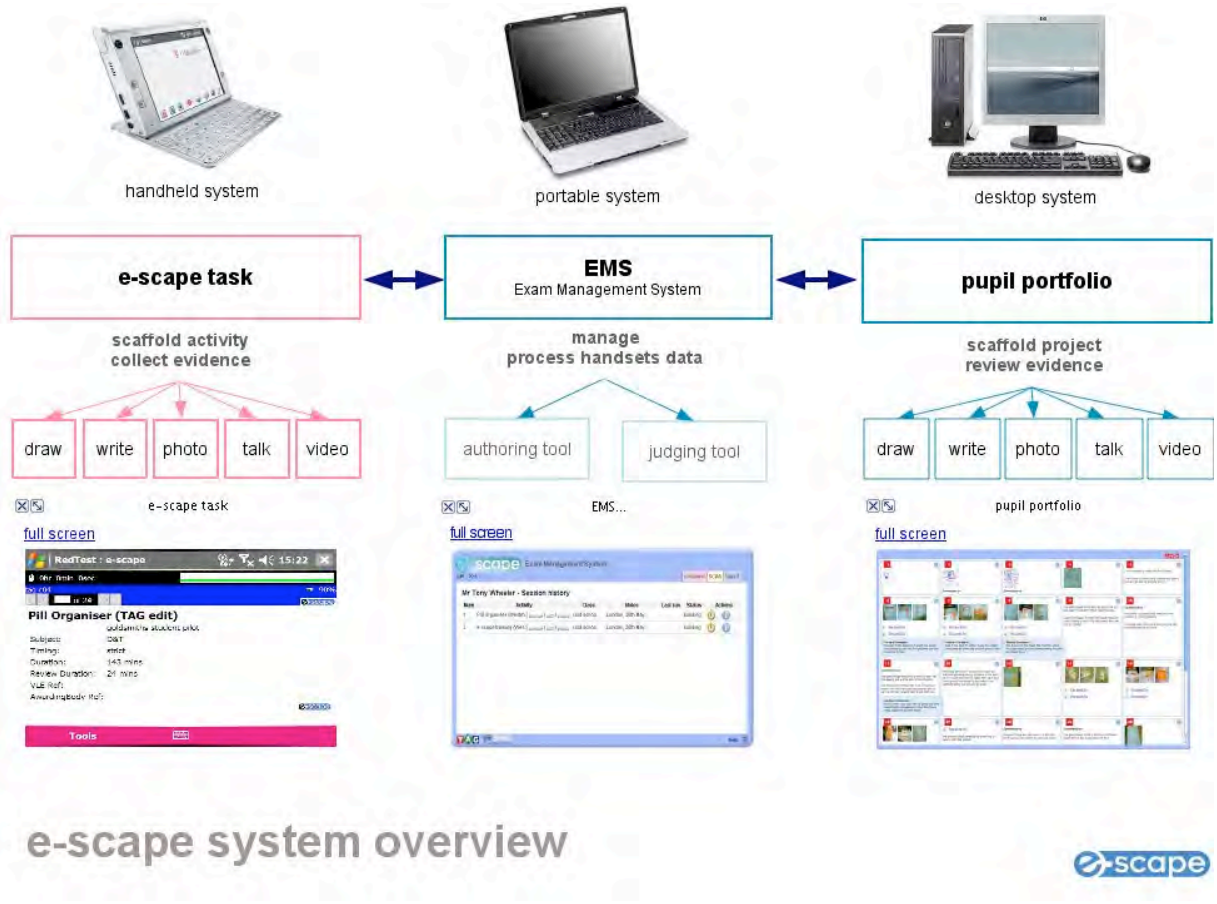
After several trials it became clear that while the 3G network could cope with sufficient data for individual use, there was an upper limit on the number of static, concurrent connections a single cell could handle. This meant that even with a good signal we could only log 10 to 12 handsets into the system from a single location.

In carrying out these trials we were pushing 3G technology beyond the experience of the mobile phone companies, where the main focus was individual rather than group use of their services.

After a number of unsuccessful trials and technical meetings we decided to concentrate on the Wi-Fi implementation that had worked successfully in the previous pilot. Further research would be needed into the possibility of utilising HSDPA in the future or applying the system to more disparate groups of pupils working at different times, or across a wider geographical area, for example on field trips, rather than in a single classroom.

system overview

The e-scape system developed for phase 3 of the project provides an integrated, end to end digital system, with 4 key modular components that pass data between them according to detailed instruction sets (APIs).



e-scape activity authoring tool

The working title for this component of the system was the SCAA tool (Structured Coursework Assessment Activity) and it allows teachers or examiners to design, author and share e-scape activities using a set of online tools accessed through a standard browser.

e-scape activity management system

The working title for this component of the system was the EMS (Exam Management System) and it allows teachers to run the e-scape activities on class sets of handheld devices.

e-scape portfolio viewer

This module allowed teachers, pupils and judges to view learners' portfolios of evidence using a standard web browser. The viewer functioned in much the same way that it had in the previous pilot with a number of refinements.

e-scape pairs engine

This was the most comprehensive technical development, digitally seaming together a number of previously disparate functions.

The development process progressed through a series of 6 overlapping stages

stage 1 the initial design document

The initial design document was created by TAG Learning in collaboration with Handheld Learning and the TERU research team. It built on the work we had completed in the previous phases of the project and presented an outline of the technical framework being proposed to deliver phase 3. The document audited the requirements of the various user groups as shown in the table and chart below.

		ROLE						
		Admin *	Principal Moderator	Judge	Examination Officer	HOD	Teacher*	Student
MANAGE	STUDENTS	X			X	X	X	
	CLASSES	X				X		
	TEACHERS	X				X		
	JUDGES	X	X					
	SCHOOLS	X						
ASSESSMENT & PORTFOLIO	RESULTS PDA	X				X	X	
	CREATE SCRIPT	X	X			X	X	
	IMPORT SCRIPT	X			X	X	X	
	DLA SCRIPT EMS	X				X	X	
	DO SCRIPT							X
	VIEW PORTFOLIO	X			X	X	X	X
JUDGING	PAIRS	X	X					
	JUDGING	X	X	X				
	CHAIR MEMBERS	X	X					

* - Researchers can view all information on all systems as "read-only"
** - Invigilators can only Run SCAT EMS

table showing the functional requirements of each user group

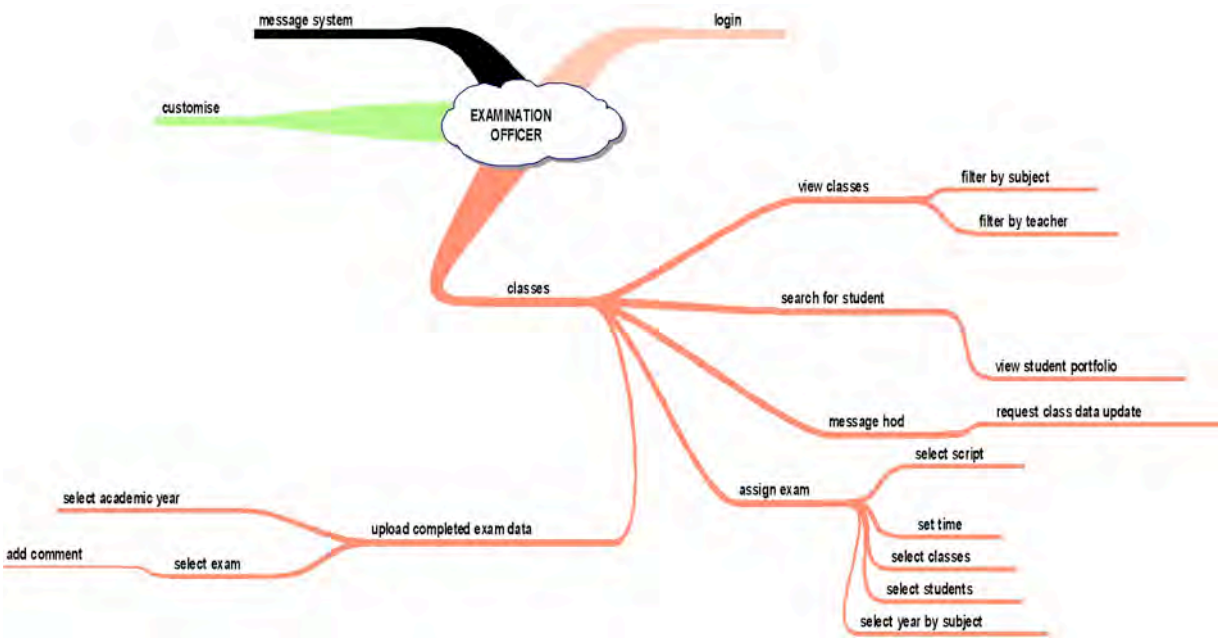


chart showing the system features available to the examination officer

Illustrated walkthroughs were also used to provide a sense of the types of interfaces that would be built and what users would be able to achieve using them.

"Box 1

The e-scape Client will automatically display the details for Box 1. The student is then able to action the Tasks within the Box. These results are automatically uploaded to the RedHalo Server, in the background, via the RedHalo data synchronisation mechanism.

As the allocated time remaining for the box counts down the student will be given a visual / audible warning when 1 minute is left to complete the current task.

The Exam Management interface will show the status of each Device and indicate that they are all waiting to progress to the next box. When the teacher has given all required instructions to the students and determined that they are ready to move on they can use the controls in the EMS to move the SCAA on to the next Box. "

In addition the initial design document described the technical issues that were anticipated in terms of coding, hardware, connectivity and interoperability, providing detailed API descriptions and XML schemas for each component.
(For the full design document see Appendix 5.)

stage 2 creating the SCAA authoring tool

This was the first component of the e-scape system to be built in phase 3 to allow researchers and teachers to create, edit and share assessment tasks. In discussion with the TERU team TAG prepared interface designs for the SCAA authoring tool which allowed the subject teams and selected teachers to create a wide range e-scape activities. The tool is available online and accessed via a password and username through a standard web browser.

Using the SCAA authoring tool

The editing tools are accessed through a standard web browser and allow teachers and examiners to create a wide range of activity frameworks.

Each e-scape task is comprised of a series of procedural sub tasks or instructions that are presented to the pupil on the handset screen. Activity authors can include text, pictures, audio, video, data, or mind maps either to offer information or set a particular task.

Having launched the authoring tool, the first step is to create and define the top-level properties of each subtask, including a title, notes and running time. Each subtask appears as a separate frame or box in the SCAA timeline at the top of the screen. Individual boxes can be dragged and dropped into different positions on the timeline and a running total of the individual and overall times of the activity is displayed and automatically updated as changes are made.

Each sub-task can be edited by clicking on it and selecting the edit button. Text images and actions can then be added to the selected box using the tools at the bottom of the virtual PDA . As you click on each of these, a new action is added to the screen display and the appropriate editing tools are displayed on the right of the PDA. The image above shows the text-editing tool.

This screen shot shows the image-editing tool. The text and image-editing tools allow teachers to create/compile SCAA instruction screens.

Pupil response components are created using the Action tool. Here teachers select a data response mode (drawing, text, photo, audio, video, data, diagram). The size or overall length of the data to be collected can be capped. It is also possible to create web-links to display a richer set of stimuli.

In combination these tools facilitate the creation of a wide and varied matrix of SCAAs with:

- different number of sub-tasks
- different times, ranging from minutes to months
- different contents and demands for each sub-task
- rigid or flexible timing
- tied to the server system or free roaming

As we have worked on the SCAA tool it has become apparent that it has significant potential to draw teachers and learners into a much more concrete debate about planning and choreographing learning activities. Already it was clear that it would be possible to differentiate the same overall activity, taking place in the same overall timeframe in many different ways simply by allocating different times for each sub-task. In this way differently timed versions of the same activity could be constructed for different sub groups or even individuals in the same class, allowing teachers to manage complex personalisation of a common task.

While in its current form the SCAA tool focuses on different media responses, it is clear that this could be usefully extended to include more pedagogic aspects of learning choreography at a future date, including components to help teachers steer the 'active' or 'reflective' skew of their interventions (do you want to get pupils thinking or doing), or the 'divergence' or 'convergence' of a learner response (do you want to get them looking more broadly or focusing more tightly).

Sharing and distributing e-scape tasks

The e-scape system provides a series of tiered activity libraries, which allow the distribution and sharing of activities between individual teachers, schools and LAs. It is possible to import pre-authored activities from one library to another so that a teacher might find an appropriate structure ready-made in an activity authored by another teacher and copy this into their own library to change the content (or vice-versa). In this way a collection of templates could be developed to help fast-forward other teachers into the process of creating their own e-scape activities, or a range of personalised/differentiated materials developed quickly for different members of a class from a common starting point.

The SCAA authoring tool now has over 100 e-scape activities in the project task bank comprising more than 20 different activity structures from primary and secondary D&T, geography, science, maths and English.

stage 3 preparing the EMS and handheld system

The EMS provides tools to set up, manage and control e-scape activity sessions. This is the hub of the system orchestrating the flow of data between all the other components.

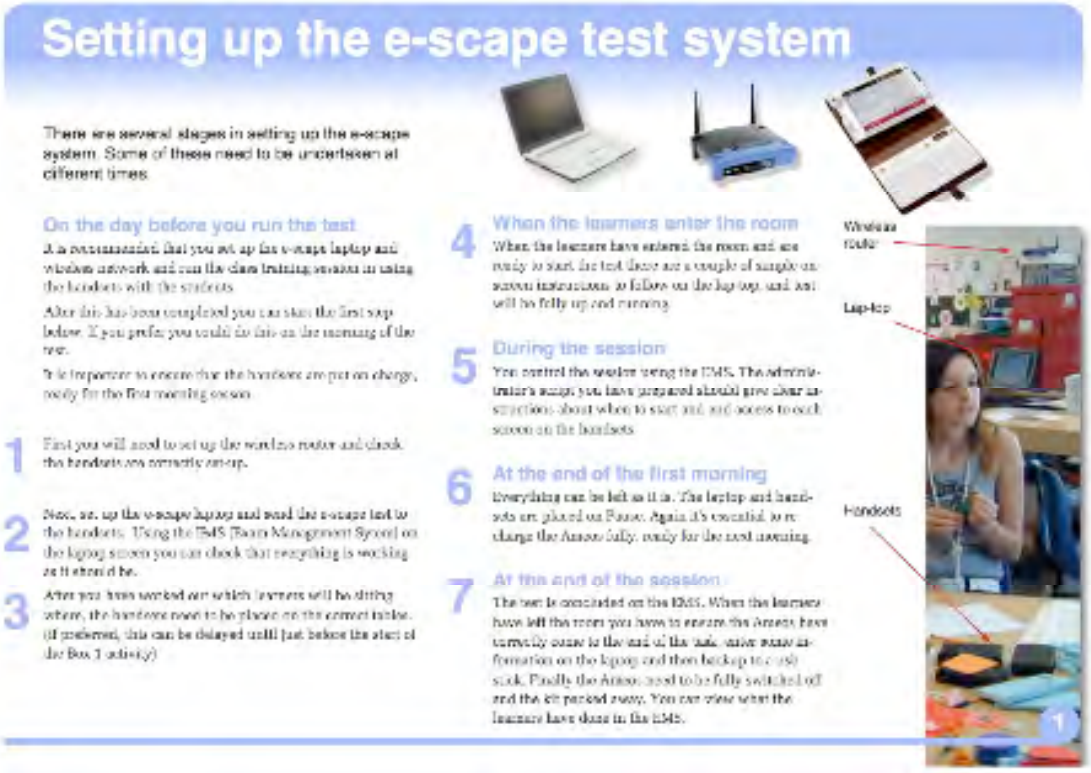
Using the Exam Management System (EMS) tool

Once a SCAA activity has been authored it is necessary to transfer it to the Exam Management System. This process is slightly different depending on whether robust 3G is available to facilitate web based delivery of the activity in the classroom. If it is, the transfer between the two systems can take place online.

However for all the main trials in phase 3 we used with a local wireless solution that required the SCAA instructions to be taken from the online authoring tool and copied into the remote server. Once the SCAA is loaded the process remains same:

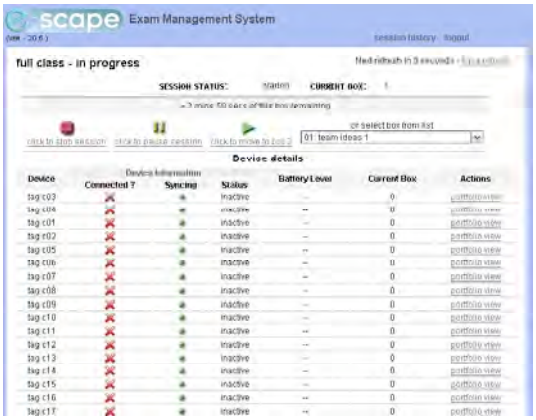
- select pre-authored SCAA
- assign to a class
- group the learners
- activate the activity
- ensure the handsets are switched on and connected to the local wi-fi network
- use the e-scape dashboard to control/run the activity

As part of our training sessions, we developed four sets of instructions that we believe are simple, transparent and foolproof when operated by teachers. One set focuses on preparation for running an e-scape activity, one for teachers who are running a pre-prepared activity, another includes additional instructions for loading SCAAs and a final set explains the use of the authoring tool. (Versions of the instruction sheets are available in Appendices 8 & 9)



Management system

Throughout beta testing and trialling a number of changes were made to the interface of the EMS. This component has become known as the “dashboard” as it provides the means by which teachers drive the activities in the classroom.



the e-scape EMS dashboard

The main aim of this stage of development was to provide a more intuitive experience for the teachers managing the activity. To help in this process we engaged the services of a 'usability' expert (Penn Smith) who attended e-scape sessions run by the team and teachers reporting not just on the screen interface but also on the training materials, script and instructions provided by the team. As a result of this process, the dashboard interface was redesigned and the support material revised paying particular attention to:

• *the use of language:* Names of items and onscreen descriptions were all revised to describe in clearer more teacherly terms precisely what they did.

• *starting the system up:* The process of initialising the system was simplified and rather than stepping through four different screens the three stages are now highlighted and controlled from a single start up screen.

• *'go next' button:* The initial e-scape EMS interface was driven using a drop down menu (a chronological list of all the subtasks in the activity). This proved very fiddly to control and as 90% of the time teachers just needed to go on to the next step, a large green 'go next' button proved a good solution. It allows teachers to step simply through the activity.

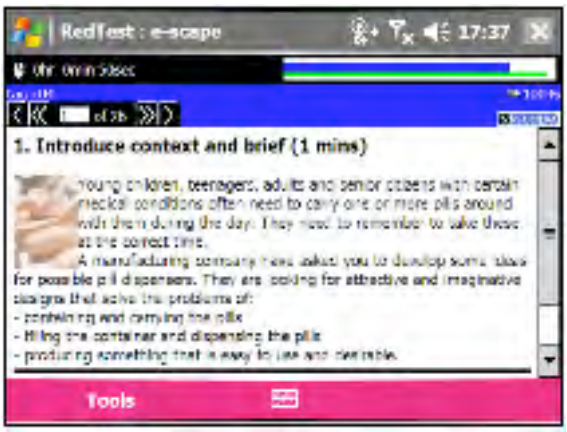
• *syncing*: The most delicate aspect of the e-scape system is the collaborative mode in which learners' work is swapped around in sub-groups to enable team support. This is a critical pedagogic support system in the e-scape activities and we have been working continually on this feature to make it more robust. Critically, when sharing and distributing work around a team it is necessary to wait for the system to upload, distribute and download the components. To give teachers a clearer picture of the progress of this file transfer, animated icons for each pupil have been included in the interface to not only show when the process is complete, but, more importantly how it is progressing.

• *'all-on-one' screen:* The layout of the dashboard screen was reorganised to show all 21 pupils in a single screen frame without needing to scroll up and down.
(For the full usability report see Appendix 6)

e-scape activity handset system

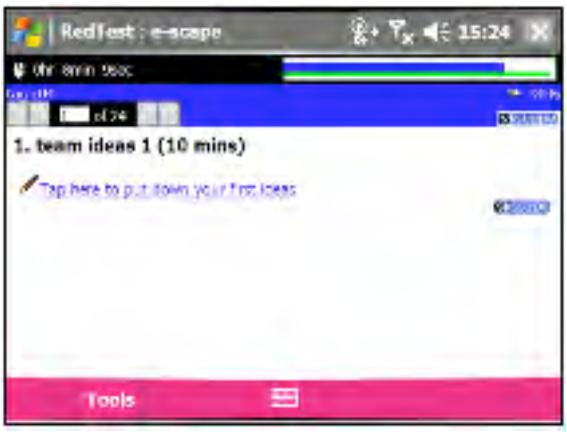
Once an e-scape activity has been created in the SCAA tool and has been allocated to a class in the EMS, the teacher/administrator can use the EMS dashboard to run the activity. The following section shows a selection of the main screens presented to learners participating the D&T Pill organiser activity on the Ameo handsets.

We determined that it was better to restrict the text displayed on any screen to the minimum, providing comprehensive instructions for each stage through the teachers spoken script. We did not judge it necessary for this trial but it would be possible using the existing authoring tools to include these instructions in text and audio form as links from each screen.



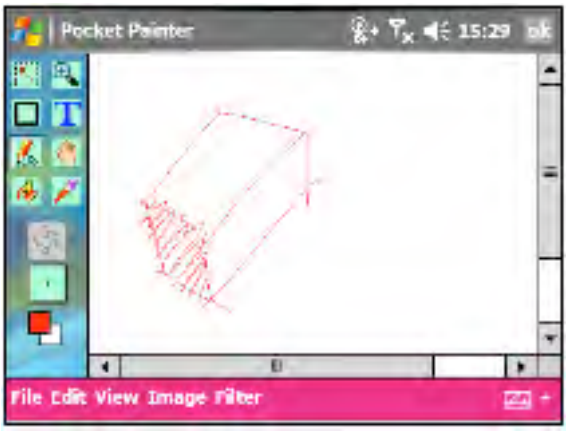
Information screens

The first screen presented to the pupils in this activity outlines the design task with text and a small image. If necessary it would be possible to include an audio version of the text as well as contextualising video clips.



Evidence collection screen

Each of the e-scape activities was administered using a common ***script*** from which the teacher read out instructions. This allowed us to keep the amount of text on each screen to a minimum. This is an example of an evidence collection screen, the teacher explains the sub task and invites pupils to click on the blue hyperlinked text that in this case takes them to the sketching tool.

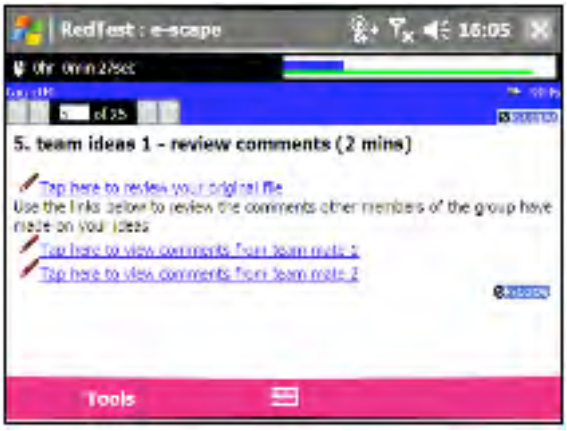


Application screen

The e-scape system links to a range of data collection tools including a still or video camera, voice recorder, text tool, spreadsheet and mind mapping, as well as the sketching tool shown here. Working with touch screen devices means that pupils can sketch directly into the handset.

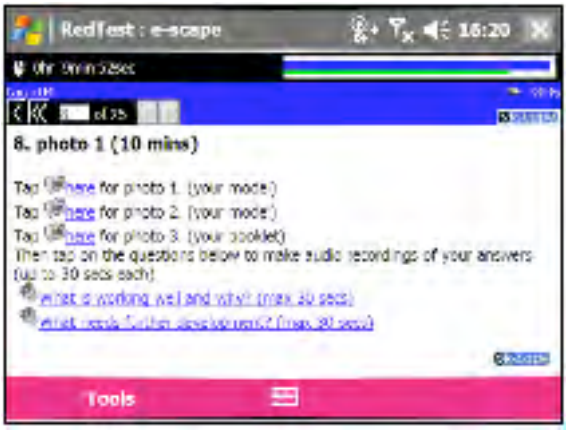
Exam mode

In the tightly constrained exam mode each of the activity screens is presented for a fixed length of time. The system sends a warning 30 seconds before the allocated time expires and at the end of the task displays the warning here. It is still possible for learners to go back and view everything they have collected in their handset portfolio, but in this strictly controlled mode it is not possible for them to edit it after the time has run out.



Collaboration

The e-scape system has been set up to facilitate collaboration. Working digitally means that evidence collected in this mode can be placed in multiple portfolios. This means that all contributions can be logged to both the provider and the receiver of ideas making it possible to provide real evidence of the contribution of individuals to their team-mates work.. The screen here shows the final stage of a 3-way sketching collaboration. We also used text based collaboration and the system can work in this way with any of the data collection tools.



Video, audio and photo capture

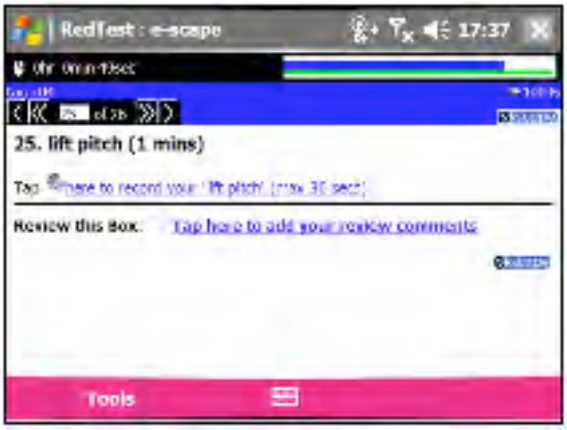
This screen shows the layout of the screen we presented after each hour of the activity. To ensure consistency the authoring tool is restricted to collecting a single data item against each link. Here we invite the pupils to take 3 photographs and record 2 audio clips.

View and Replace

It was important to allow the learners to review their photos, video and audio, and re-record if they were not happy with their first attempt. Clicking back into a collection link once an item had been saved displayed this screen. “Record new” created a new file to replace the original, “view exiting” displayed the saved file. Additionally, if the item was a photograph, the system launched this using the sketching software, allowing pupils to annotate and edit their images.

Text entry screen

The Windows mobile system offers a range of text entry tools including real and onscreen keyboards as well as transcribing handwritten text into type dynamically on the screen. In the training session we allowed pupils to experiment to discover which was the most effective method of text entry for their writing style and experience.



Fast back review screens

At the end of the test session it is possible to set the e-scape system to open a comment box on each (or a selection of) sub task(s). The screen shows the additional link that then appeared at the bottom of the screen. Pupils could review, but not alter, the evidence they had collected and then add a note to explain anything significant that was either not apparent from the data, or reflections on what they would have done differently.

We produced a training power-point presentation with screenshots of each of the key interface features that the teachers used in a familiarisation session with the pupils immediately before they ran the tests. This can be found in Appendix 10.

stage 4 refining the portfolio display

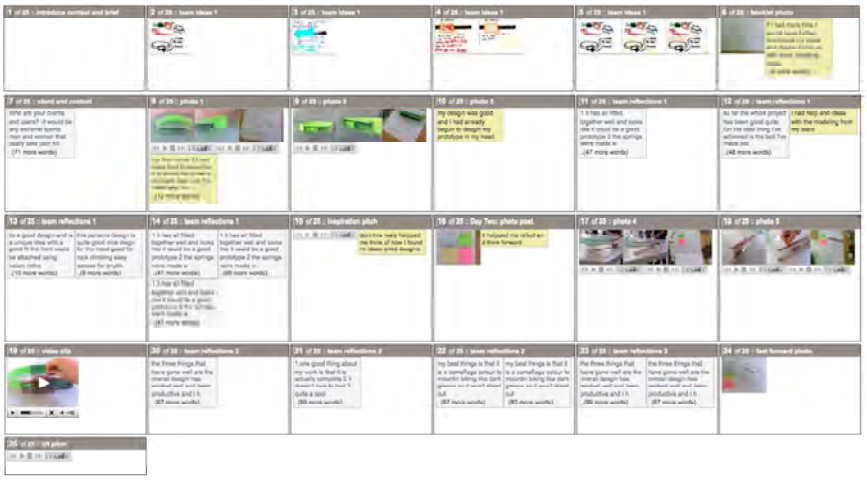
Portfolio access system

The online portfolio access system presents the evidence trail of individual student performance collected dynamically during e-scape activities. The system is based on a number of key principles.

“blink-ware”

In order to mirror our experiences of working holistically with paper portfolios, we made sure that it was possible to present an overview of the entire contents of each portfolio in a single screen. We have seen this approach referred to as “blink-ware” and has involved the consideration of how to “thumbnail” each element to give an instant snapshot of the whole portfolio on a single screen.

layout of an e-scape portfolio overview from phase 2



layout of an e-scape portfolio overview from phase 3

“zooming” in and out

In order to access the detail of the portfolio from the overview screen, each box acts as a button. Clicking on an item enlarges the media elements and allows the contents to be examined in detail. For the D&T activity these included:

sketches

Students sketched their initial ideas in the first 5 minutes of the D&T activity using the paint application on the handset. The system passed these sketches around the group of three so each could add comments and other ideas to their team-mates work. The portfolio shows each learner’s initial ideas as well as their contributions to their two team-mates and all the comments they received.

photos

After each 40 min modelling phase learners took 3 photos of their work using the handset. The developers included a further zoom, contrast controls and rotate features to ensure judges were able to gain maximum benefit from images some of which were less clear than others.

sound files

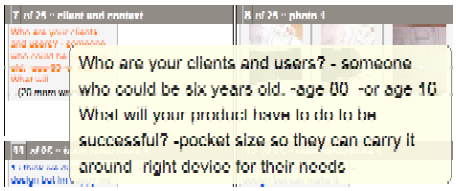
Judges could listen to the sound files recorded at each stage of the activity using standard audio controls

video

As the speed and storage capabilities of the technology have advanced since the last pilot, it was possible to work with video clips in this version of the system. These appeared in the portfolio and could be viewed in thumbnail or full screen mode

text

The portfolio overview shows the beginning of each block of text entered by the student. To speed up the viewing process this was set to expand and display the full text entry when the judge hovered the cursor over the box.



The original phase 2 portfolio contained 20 frames of data, in phase 3 this was expended to 25 frames as we introduced a new audio recording of each learner's "inspiration pitch" to help judges see where each idea had come from. We also reorganised the display of the team feedback sessions.

future e-scape portfolio features

All judges reported that there were some critical elements of the portfolio that, if they were missing, made it more difficult to make a confident judgement about the overall performance. In particular the video, first ideas sketches and the 'fast forward' box at the end. In many cases it was difficult to determine whether the lack of evidence was due to technical problems or that the students had decided not to capture and post that evidence.

In future we have suggested that it may be necessary to develop a feedback system to alert the teacher running the e-scape activity that data is missing at the time it is collected and offer options to capture and mark that data after the timed sequence. Discussions with the development team suggest that this is entirely do-able.

stage 5 pairs engine

e-scape pairs judging system

In phase 2 of e-scape, the paired assessment was completed with a very rudimentary system. Judges were provided with a list of portfolio numbers printed on paper. They accessed each portfolio online by searching through the school and class lists. When they had manually found and opened both portfolios on screen they made their judgement and noted the winner on the paper list in pen or pencil. On completion of the judging, the lists were sent back to the research unit, compiled and sent on to Alistair Pollitt who ran the pairs comparison on his local system.

Considerable development has gone into digitising this whole process so that we now have a secure system that manages the judging activity online from end to end.

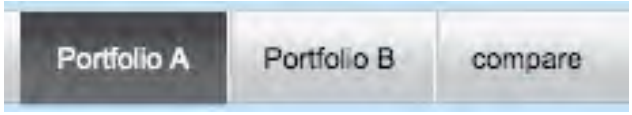
A nominated administrator (we imagine the principal moderator at an awarding body) is provided with a secure log-in that presents a suite of tools to manage the process, including allocating centres to be assessed, judges to make the assessments and examinations/tests to be assessed.

Once an assessment session has been created using these tools, the system automatically sets up accounts for each of the judges, calculates the overall number of judgements that need to be made and then dynamically allocates pairs of portfolios and sends these to each judge.

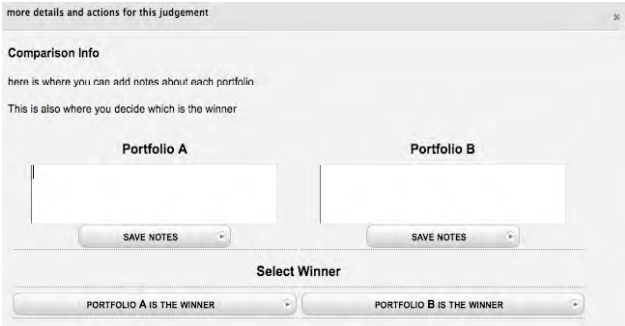
Session History for Alan Beech

Spec	Description	Status	Last update	Global % complete	My Running total	Action
escape phase 3, Design and Technology	main phase 3 design and technology judging session	running	9th Nov '08 23:06	88.18	63 of ~126 (50%)	start next judgement

When a judge logs in they are shown a list of assessment sessions (they may be involved in more than one). When they click on an active session they are taken to a screen that shows how many judgements they have made and how many are still to be made. Clicking on the “start next judgement” button automatically selects the next comparison the system needs and presents these two portfolios to the judge.

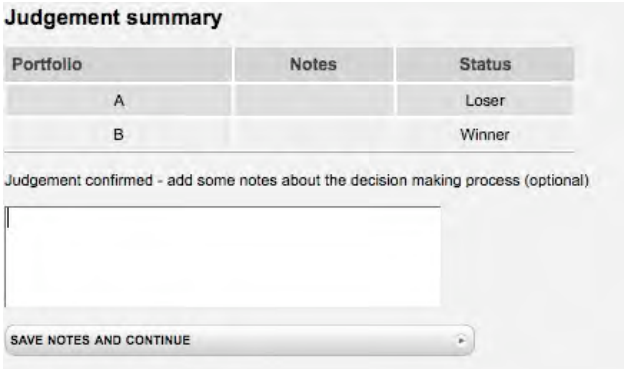


The judge can toggle between the two portfolios to be compared using the buttons in the top menu bar.



Once they have completed their comparison and are confident they have identified which is the better, they click the compare button. Here they can then add notes about each portfolio and select the winner.

The system includes additional confirmation steps to minimise mistakes in selecting the winner.



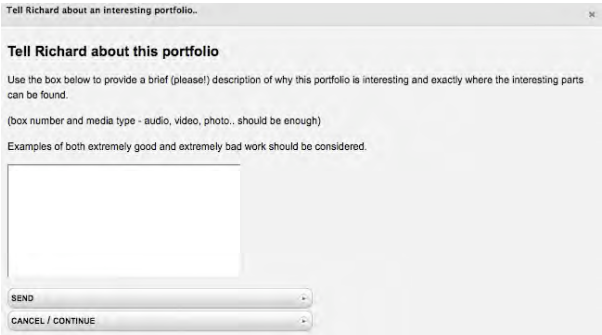
When a winning portfolio has been selected the judge is invited to make notes on the comparison process, i.e. how difficult it was, what features of performance swayed their decision.

Judgement History

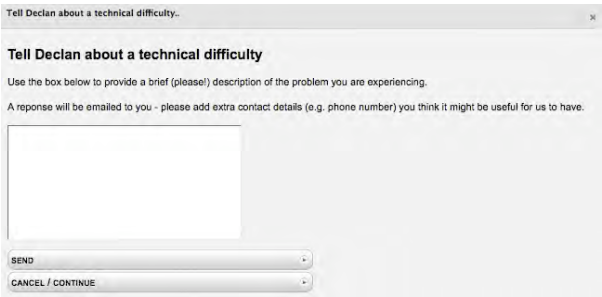
Start time	End time	Duration (hours:mins:secs)	Portfolio A	Portfolio B	Judgement notes	Winner
31 Oct '08 09:41	31 Oct '08 09:45	00:03:27	lost in basic tilty model but lots of ideas noted on booklet and fast forward (503)	basic pill sucking system (598)		Portfolio A
31 Oct '08 09:33	31 Oct '08 09:41	00:08:36	whacky twisty pill sucker (547)	lost in basic tilty model but lots of ideas noted on booklet and fast forward (503)	I found this one really difficult, which is better 1 big whacky off the wall idea or lots of little alternative rooted ideas? Not sure I made the right decision, I think I may be swayed by the kids who seem to be taking it more seriously	Portfolio B

The system collects a range of data about each judging session that is available to each of the judges. This includes the time taken to make each judgement, together with the judge's notes on each portfolio, the judging process and the winner of the comparison.

Other new features



Tell Richard
This feature was added so that judges could inform the research team of particularly interesting or unusual portfolios (or even individual boxes in them).



Tell Declan
This feature was added so that judges could inform the development team of any technical difficulties they encountered.

All judges agreed that it would be helpful to be able to open multiple portfolios on screen as this would speed up the comparison process. We have also discussed the possibility of developing a tool to reorganise the individual portfolio frames/boxes on screen to allow judges to focus on the key discriminating evidence. Another suggestion has been to investigate restricting the number of judgements an individual judge can make ahead of the rest of the assessment team, to avoid skewing the outcomes at particular stages of the process.

e-scape comparative pairs system administration tools

The e-scape pairs system provides a number of management tools that allows the system administrator to set up, monitor and then analyse the out-turn from each assessment session.

The main session screen allows the system administrator to view a list of all the pairs assessments, it indicates those that are currently active, shows a history of completed sessions and provides a facility to set up new sessions.

escape phase 3, Design and Technology

main phase 3 design and technology judging session

Judges	Portfolios	Status	Last update	Show stats	Action
28	352	running	9th Nov '08 23:06	17 rounds complete ~ 63 judgements until next estimation round 3105 of 3520 judgements complete (88.21%) parameter graph	END PAUSE ABANDON

Running

Session screen for the D&T pilot judging session

From the individual session screens the administrator can view data on the judges working on the assessment. The main view shows overall progress and a measure of the ongoing consensuality of each judges' comparisons. Clicking on a judge's name in the list shows the detailed judgement history of every comparison made by that judge.

Judge (click for history)	Progress	Extra	Misfit figure	Active status
tony wheeler	127 of ~126 (100.79%)	0	wms=0.81881684451527 -1.5843209306379	De-activate
richard kimbell	127 of ~126 (100.79%)	0	wms=1.2171667702992 1.4879882836919	De-activate

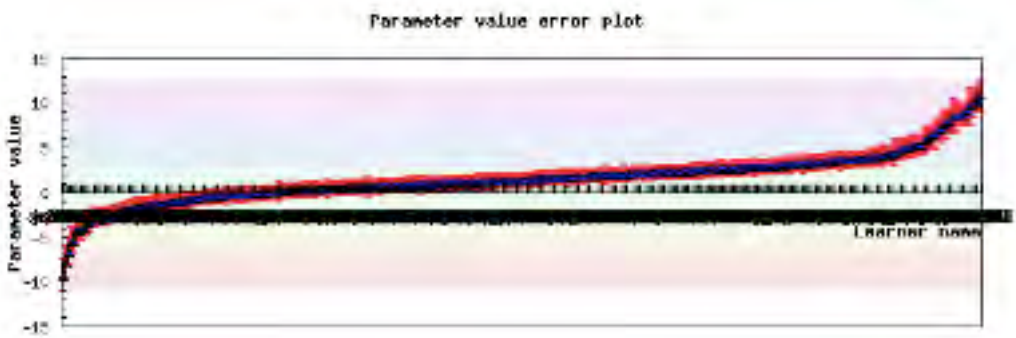
Judges' information screen

The system administrator can also access detailed information about the judgement history for each portfolio from the main session screen by clicking on the number of portfolios.

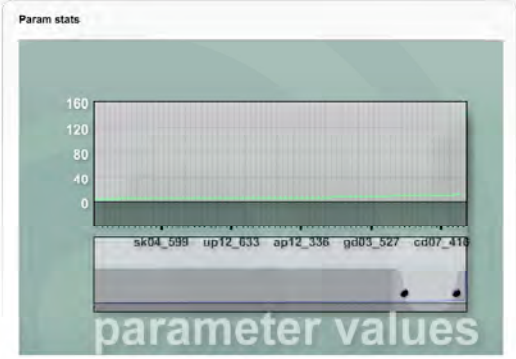
Portfolio Judgement History

Judgement Time	Judge	Link	Paramter	Compared to	Paramter	Notes	Winner?
10th Nov '08 00:25	Manjinder Sangha	ad09 student(305)	0.272546	up13 student(629)	0.27544		NO
10th Nov '08 00:22	Manjinder Sangha	ad09 student(305)	0.272546	nc10 student(570)	0.110176		NO
9th Nov '08 15:02	Andy Millicheap	ad09 student(305) box with stars on	0.365796	bb20 student(388) us mailbox dispenser - thinking bout mechanism	0.431134		NO
9th Nov '08 14:59	Andy Millicheap	ad09 student(305) box with stars on	0.365796	pr11 student(582) velcro stick on	0.349591		NO

A full history of all the comparisons made and notes taken by each judge for each comparison can be displayed by clicking on each portfolio.



At the end of each round of comparisons the system automatically calculates the position of each portfolio based on a parameter value derived by calculating the probability of the portfolio winning or losing each comparison. Graphs of these calculations are available to the system administrator throughout the judgement process.



Once the judgement process is complete the positions of the portfolios are transferred to an interactive display where it is possible to open each portfolio by clicking on its marker on the final parameter graph.

For the D&T, geography and science trials the engine was set up to operate through 20 'rounds' of judging. A round is when each portfolio is compared with one other. At the end of 20 rounds (each portfolio compared to 20 others) the data is so stable that additional comparisons make no difference to the rank order and only very minimal differences to the statistics. The law of diminishing returns operates and it is very unlikely that in a real situation the judging would need to go through more than 20 rounds. In fact in the judging process just completed, the data was stable (in the sense that the rank order remained unchanged) after only 12 rounds.

The first 6 rounds of judging are used to create an approximate rank order - and this is based on a 'Swiss tournament' methodology. All the portfolios start randomly and each is compared with one other - which produced 50% winners and 50% losers. In round 2, pairs of winners are compared and pairs of losers. This results in some having won twice, some having won once, and some having won not at all. In round 3, these groups are again compared internally (i.e. a twice winner with another twice winner) and so on. At the end of round 6 the whole sample has been divided into seven clumps of portfolios. In the top clump are those that have won all 6 judgements, and in the bottom clump are those portfolios that have not won any. There is no internal order to these clumps, but there is a good chance that all the work in clump 1 is better than all of that in clump 7.

At that point - from round 7 onwards - the full estimation algorithm gets to work and generates a complete rank order. It is at this point that the ***presentation*** of portfolios to judges changes somewhat. In the 'Swiss tournament' rounds, each judgement is based on a new pair of portfolios being presented. But from round 7 the portfolios are 'chained' to make the judging process easier and quicker. With chained pairs, if (e.g.) I am judging portfolios 66 and 184, when I decide that one is the winner (e.g. 184) then when the engine moves to the next pairing it retains one of the previous pair and finds one other new one for me to compare it with. This means that for every paired judgement, the judge only has to get to know one new portfolio. All judges reported how much easier and quicker it is to operate in this chained mode.

More detailed information regarding the statistical principles and aspects of the specific algorithms used to drive the comparative analysis can be found in Appendix 11.

Some emerging technical difficulties through the summer testing schedule

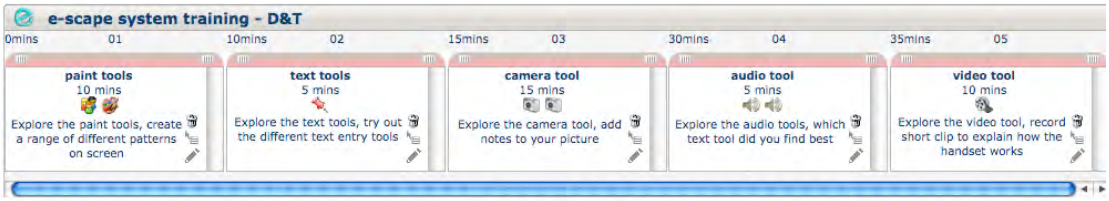
A number of technical problems with elements of the e-scape system became evident during the trials and the summer testing, most of which were known before we embarked on the full testing programme in June/July 2008. They have been the focus of continuing 'fixing' by our technology partners.

• *running tests in 3G mode*: the problem of multiple connectivity from a fixed location.
• *sketching and text collaboration*: some problems arose with these tools freezing during the collaboration cycle. The issues was resolved in the latest version of the system but was not available for the devices already in circulation (in hot schools).
• *review mode at the end of timed tests: W*e had some difficulty with the timings on this aspect of the system. Again they were fixed and available in the latest version.
• *pupils reviewing their work during the activity*: It has always been possible for teachers to review all pupils' work as it unfolds - by using the portfolio display tool on the EMS dash-board. But it was only possible for learners to review their own work on the device when it was active on a sub task. It is now possible for pupils to look back at everything they have done, viewing (but not editing) the material.
• *locking out other applications:* It remains a problem that learners' interest and enthusiasm for the hand-held devices results in them exploring the range of other applications and features of the handsets. We understand this urge, but since all programmes take up working memory, this occasionally causes system overload and freezing. Whilst it is possible to remove other programmes, it is not simple, requiring a bespoke build of Windows mobile.
• *bugs in 3rd party applications:* We are aware of a couple of bugs in 3rd party applications, e.g. the text tool in Pocket Painter crashes the handset. While we warned pupils against using these features, the instructions were not always followed.

5. creating and trialing SCAAs

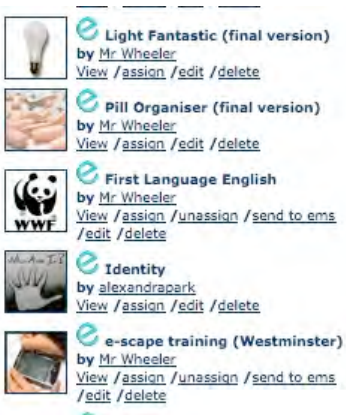
the SCAA authoring tool

The SCAA authoring tool was the first functional piece of e-scape software to be completed in phase 3. The TERU team required the tool as soon as possible to experience creating D&T activities themselves in order to prepare for and train the subject teams and teachers to develop their own.



Familiarisation activity subtasks

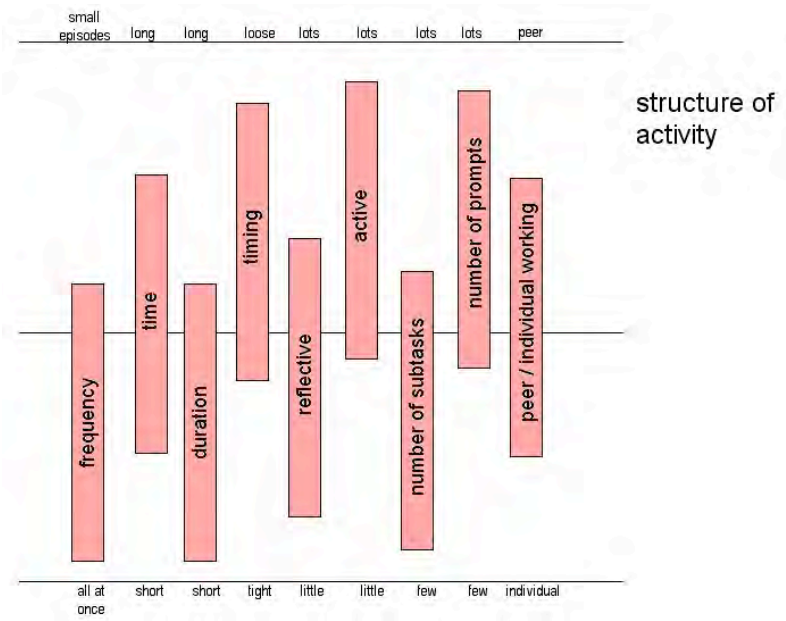
As soon as the software was in a functional state the central team began to develop activities, that, as well as familiarising us with the SCAA tool and exploring its potential, helped to identify problems with the system itself.



In this period we created a series of activity structures, in particular one for the Light Fantastic activity we had used in the initial phase of the 'Innovating assessment' project, a shorter 60 minute familiarisation activity structure, that each of the team took a copy of and personalised, and several versions of the Pill Organiser activity.

As well as refining the interface and the functionality of the SCAA tool, we also analysed and modified the format and contents of the activity structure. It was apparent that there is significant potential to draw teachers and learners into a much more concrete debate about planning and choreographing learning activities. As we have outlined above, it would - for example - be possible to differentiate the same overall activity, taking place in the same overall timeframe in many different ways simply by allocating different times for each sub-task. In this way differently timed versions of the same activity could be constructed for different sub groups or even individuals in the same class allowing teachers to manage complex personalisation of a common task.

Now that we have a concrete example of how a tool such as this might work, it has been possible to speculate on a wide range of other prompts that might also reside within the system.



Altogether over 100 SCAAs have been created and these can be viewed in the e-scape system task bank. On reviewing these developing SCAAs we realised that for the high stakes subject tests (as opposed to curriculum developments) it would be necessary to agree a more standardised set of protocols for describing and presenting actions within the system.

Teacher instructions:

Our initial inclination was to pack lots of instructions into each PDA screen. However this proved not particularly helpful to the pupils. In the previous version of the e-scape task we had kept the screen instructions to a minimum and relied on the administrator to speak these from a script. Lots of text on the screen is confusing and difficult to read and we agreed to restrict this to the minimum to explain the PDA function rather than the whole context of the box, which is better communicated through the spoken instructions from the administrators' script.

As a result of this, TAG revised the system to allow for a teacher notes/instruction box, which helped to build the activity script, and is only visible to the teacher on the server control screen.

In the previous paper and digital versions of the d&t activity we were careful to make the process 'task-agnostic'. Only the activity description on the cover of the workbook changed while the format and procedural prompts were the same for each activity. We applied the same rules in the digital versions of the D&T, geography and science tasks.

Common interface/layout/commands:

In their early drafts the subject teams used a range of different ways to call up the PDA tools. We agreed that we needed to standardise a common format to call up tools as we did in phase 2, for example:

"Tap **here** to..." (with the action link on "**here**")

Where questions required a typed answer the format should be (with the action on the whole of the highlighted question):

"Tap on a question to answer it.

What is your interpretation of your results?

Where questions required an audio answer the format should be:

"Tap on a question to make audio recordings of your answers (up to "X" seconds each)

What is happening in the photo?

A more detailed explanation of the SCAA tool can be found in section 5 and an overview of the whole system is provided in the leaflet in Appendix 04

New authoring features for phase 4

A number of refinements were incorporated as a result of suggestions by the various teams working with the authoring tool, for example:

- Greyed out boxes (when they are checked as inactive)
- Drag and drop (to reorder individual boxes in the timeline)

We also compiled a list of refinements that would be valuable in the next refinement of the system including:

- 'Copy' and 'paste' features for individual boxes
- Ability to group boxes
- Ability to create and combine modules from groups of boxes.

the design & technology SCAA

The main focus of the D&T team at this stage was to develop a new e-scape version of the pill-dispenser activity. This was the same activity that featured in the 2006 pilot and it was decided that it should also be the core task for the 2008 d&t pilot. The main reason for this decision was that this element of the research design was all about *scalability*. Could we enable teachers to run the activity for themselves rather than having us (the research team) delivering it for them as we did in 2006? Having two directly comparable data sets (one from 2006 and one from 2008) would therefore give us some purchase on the question of whether learners' performance was affected by the change of responsibility from the research team to teachers.

We made some minor modifications to this version of the activity to take advantage of the features of the new system, in particular introducing the use of video. In phase 2 the capacity of the available network technology had not been sufficient to cope with multiple video transmissions. The new system proved to be able to cope with a class set of short (30 sec) video clips which allowed pupils to describe their ideas more dynamically and we replaced the final photographic session with a video presentation.

The only other significant change was the addition of an audio "inspiration pitch" at the end of the first day to help assessors identify where pupils' initial ideas had come from See section 11 for more information about the inspiration pitches. A full description of the activity is available in Section 2.6 of the final report of phase 2 of the project. (Kimbell et al 2007 p 41)

For a more comprehensive picture of the Pill organiser activity we have included:

- preparing for e-scape leaflet Appendix 12
- d&t pill activity overview Appendix 13
- the administrators script in Appendix 14
- the pupil work book in Appendix 15
- user profile cards Appendix 16

There is also a Teachers TV video of a trial of the phase 2 version of the task at Saltash School in Cornwall online at: http://www.teachers.tv/video/3306

Since the activity itself was very fully tested in 2006, we did not need to establish anything else about it in terms of it being an authentic valid, d&t activity. We did however need to be confident that it ran at least as well on the new system as it did on the 2006 system. To that end we scheduled a series of largely *technical* trials in school settings with full groups of 21 learners.

science and geography SCAA development training

We worked together with both subject teams to help prepare them to develop e-scape geography and science activities. The process we employed was largely experiential and provided a framework for subsequent training of the 'hot' school teachers who would be developing their own curriculum activities using the SCAA tool.

There are a number of key issues that underpin the approach to performance assessment that has been developed at TERU over the past 25 years. Identifying the critical aspects, and then sharing and exploring these in a single session is a serious challenge. We used the following programme with the two teams in Rugby in June 2007.

1. starting with the QCA importance statements

The principles of holistic assessment are central to the work at TERU and the question of what overall design capability looks like should be familiar to many d&t teachers. The notion of holistic performance in science and geography does not seem to be so commonly recognised. To provide a platform for developing a common understanding of holistic assessment, we initially worked with the QCA importance statements as overall performance descriptors. Each team then evolved a more personalised interpretation of their statement. See Appendix 17 for the QCA statements and Appendices 18 & 19 for the geography and science final reports.

2. explaining the e-scape structure, sequence and instructions

Having established that we wanted to find some way of identifying who was the best geographer/scientist/designer, we explored the basic components of the TERU e-scape d&t tests. We viewed the Saltash video and explored the main techniques we employed such as the teachers script, short bursts of activity, modelling and handling collections and modelling kits. We prepared a leaflet on setting up these components for e-scape assessments (see Appendix 20).

3. running through a system familiarisation activity

We then ran a shortened version of the pupils familiarisation activity to show the range of media collection facilities available on the digital system and to give the teams a feel for the pupils experience when taking an e-scape 'test'.

4. de-constructing paper based d&t activities

A key aspect of e-scape assessment is the tight choreography of a series of sub-tasks. To illustrate how this technique could be employed we had prepared a number of de-construction activities where we analysed the interventions in existing d&t e-scape activities, logging both their pedagogic intention and the difference between paper based and digital responses. (See Appendix 21 for an example of a completed activity analysis chart)

5. exploring ideas for subject based activities
We worked with the subject teams individually to explore their ideas for e-scape type activities in their own subjects

6. breaking activities down into sub tasks
We worked together to start to break the overall activities into sub-tasks and consider what sort of interventions would help to both stimulate and steer the activity as well as providing a stream of evidence to help make a judgement of their overall capability. (See Appendix 22 for an example of a blank activity analysis chart)

the geography SCAA

The geography team was based in Bath Spa University, with most of the e-scape development work operating in school settings in South Wales.

The geographers developed their paper-based tasks into the authoring tool over the 2007 Christmas period and we went through them at the training session on 25th February 2008. The team co-opted a teacher in Pencoed school near Porthcawl, who had good working relationships with the geography teachers in Dyffryn school close by, near Port Talbot. The trials of the e-scape geography activities took place in what was – in effect - a cluster of schools.

The geography team developed a range of tasks: rainforest / the quarry / Cheddar gorge / Porthcawl etc. which represent a rich diversity of experience in geography. Their approach to e-scaping them was to take short parts of the activities and run them in ordinary lesson times. The whole activities were then assembled and run. This seemed to us to be a perfectly sensible approach, particularly as it allowed the technology also to be grappled with a step at a time.

The incremental approach adopted by Fred Martin and Tom Biebrach (the co-opted teacher) was supplemented by another element in their development process. The problem was that (unlike other e-scape activities in d&t or science) the local setting is important. It does not make much sense to run a Cheddar Gorge activity in Huddersfield. Obviously there is a world geography context (e.g. rainforests), but one of the features of geography that can be enriched by e-scape is live fieldwork in local settings. The 're-branding Porthcawl' activity illustrates this potential.

The challenge then was whether it was possible to make such local activities into national tests. The geography team developed an approach that invoked creating a template for a generic activity but one that could recognise and adopt the local setting.

> *This afternoon, I had a go at a generic planning-decision activity. I have been interested in creating a generic template that a teacher could use to look at* ***any*** *local issue, supplementing it with their own local resources (or letting the pupils find local resources). I know the brief is to trial one item in June, but I am finding it far more productive to explore several at the moment.*
> (Fred Martin e-mail to e-scape team 4/3/2008) [our emphasis in bold]

This seemed to us to be an appropriate and positive line of development given the particular concerns in geography. It also contributed substantially to taking forward the geography discipline, and particularly in the tricky area of fieldwork and direct data

collection off the school site. The mobile facility – and yet under the teachers' control - opened up all sorts of potential for that was later to be explored in the geography e-scape task. The full geography report is available in Appendix 18.

the science SCAA

The science e-scape team comprises Professor Dan Davies (Bath Spa University) and Gareth Price (Sheffield Hallam University). Their approach to the task was somewhat different to the geography team. At a meeting at the outset of the project two task areas were identified within the science curriculum and these were pursued semi-independently by the two members.

The two tasks in science were:
• road safety … the learners acting as advisers to the local council and aiming to make recommendations about traffic calming measures outside the school. The task involves learners designing and conducting experiments in acceleration, deceleration, and impact damage, interpreting their data and then making scientific sense of official data from the Dept of Transport about the effectiveness of a number of traffic calming systems.

• reaction timing … examining the effects of a number of factors on the speed of reaction of a driver.

The first task has a broadly physics bias and the second one has broadly a biology bias, though there is a strong component of 'science in society' attached to both.

The road safety task in particular was trialled in schools 6 times – and was systematically developed into a coherent and convincing piece of science. The development process underlined some of the issues that we have been aware of in design & technology for some years.

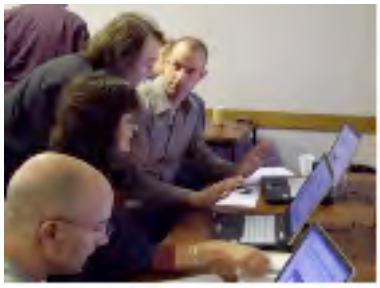
The first key issue is time. The task development process was closely attuned to trials in schools, and specifically in the context of one-hour science lessons. So the activity – which is intended as a complete 3 hr task – was trialled in three bits and then put together and trialled again ***as a whole*** over a 3 hr morning, and the added-value effect was remarkable. We are strongly of the view that whilst schools may be inconvenienced by having to create whole-morning sessions to run the activity, they will be fully rewarded by the result.

A second key issue concerns resources. The developers were anxious to ensure that schools were not inconvenienced by having to provide extensive resources to run the activity – so it was all taken in especially for the purpose. The value of 'handling collections' and 'modelling' time is well understood in design & technology, but is not routinely to be found in science lessons. But having explored these features of the d&t activity, the science developers adopted them and were enthusiastic about their effects. The full science report is available in Appendix 19.

6. preparing for the 2008 testing

training sessions

We have explained above (section 3) that we were working with two groups of schools. The first group we referred to as the 'hot' schools, that were essentially development partners with us. We met this group early on while the software was still under development and as with the subject groups above provided training in the authoring tool so teachers could develop and explore the curriculum potential of the system. They later joined the second group of 'cold' school teachers for training in how to administer the tests.



• hot schools training

The hot schools teachers all met in Birmingham University on the 8th/9th January 2008 for training in the use of the authoring tool. The idea was that if we trained them to use the authoring tool then, they would have about 6 weeks in which to develop their activity ideas before the Ameos were available for them to run their own e-scape-like activities after February half term. We chose Birmingham for its central location and the university provided a good venue for what proved to be a terrific experience.

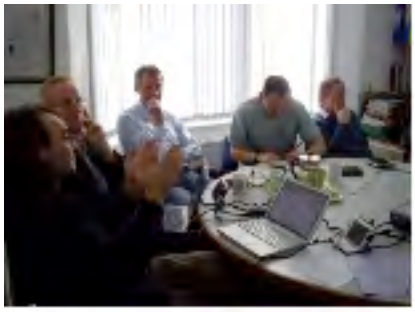
The 6 schools each brought two teachers – so that once back in their schools they had the opportunity to work together on activities. The two days involved some whole-group debate; some individual tuition on the system; some school-pair idea development work; and some presentation of ideas back to the whole group. See Appendix 23 for the agenda for the two-day training session.

Towards the end of the session, we agreed dates for delivering the kit to schools after half term – mainly during the weeks of 25th Feb and 3rd March 2008.

geography and science team system training

In addition to the sessions focussing on SCAA activity development we held a 'system-training' day at TERU (25th Feb 2008) for the subject teams. We had to teach them how to get their activities out of the authoring tool, load them into the Exam Management System (EMS) and make them active for learners on the hand-held devices. We started the process by working through these prospective SCAAs, and spent the morning reviewing them in relation to a number of key pedagogic factors, for example :

• the role of screen-based instruction
• the assigning of resources (pictures / text / video etc)
• the sequence of sub-tasks (e.g. in relation to active/reflective)
• the time-scale of sub-tasks



This had the effect of enabling the teams to see how they might refine their SCAAs. Thereafter we introduced the EMS and showed the teams how to get to the point of running the SCAA as a real classroom activity.

In the light of technical questions over the 3G signal band-width a decision was made that at this point we would prioritise the use of the wi-fi version of the system. This version of the system was therefore introduced in terms of:
• outlining the structure and operation of the exam management system (EMS)
• identifying and nominating groups of learners and assigning the task
• running the task with the Ameo handsets.

At the completion of the training session, we agreed dates for delivering the system to the geography and science teams so that their trialling in schools could be commenced.

TAG produced a guide to using the SCAA tool (see Appendix 24) and we provided a guide to setting up and running the e-scape system (see Appendix 8).

DIY kit for e-scape system training

Our starting point for training teachers was to think back through the 2006 (phase 2) testing and attempt to recall exactly what we had needed to do to make those tests successful. We concluded that several categories of things needed to be understood by teachers before they could be expected reliably to re-create this success on their own. The categories were

i) Choosing and setting up the studio/workshop
ii) Running short/sharp e-scape design tasks
iii) Creating handling collections to contextualise the task
iv) Creating a modelling kit to enable learners to model prototypes
v) Setting up the e-scape system
vi) Running a familiarisation session for learners to get to grips with the handheld devices
vii) Doing a ‘walk-through’ of the activity

The materials we developed to support teachers through these processes were known as the DIY kit, helping teachers to do for themselves what (in 2006) had been done by the central research team.

In transforming this list into a training schedule, we were faced by a paradox. Some of these things required an extended time in advance of the testing to enable teachers to compile the resources (e.g. handling collections), but other things (e.g. setting-up of the

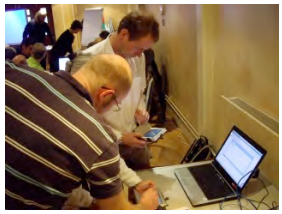
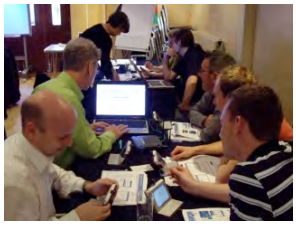
e-scape system) needed to be as close as possible to the test date so that it was fresh in the minds of teachers when they were required to run it.

This led us to the conclusion that we needed two training days
Day 1 - in regional clusters - dealing with items i), ii), iii) and iv) in the above list. One training day was established for each of the 4 regional groupings (SW, NE, W Midlands, London) in the weeks of 21st and 28th April.

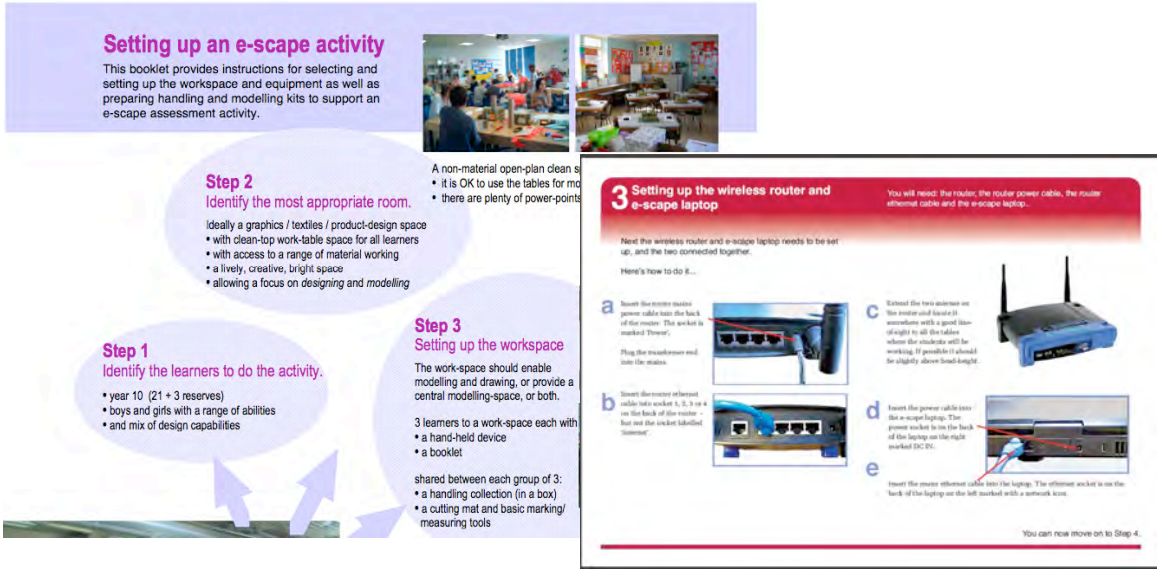
Day 2 - in central London – dealing with the technical demands of setting up and running the e-scape system. Since the testing was due to start immediately after the summer half term, a single day just before the half-term (20th May) was identified for ALL the teachers involved in the testing to come to London and experience a full day of technical training dealing with items v), vi) and vii).

Having settled on and agreed these dates with all the schools, our immediate task was to create the training materials that could be used at these sessions – and that teachers could take away with them as support for their activities.

The core resources exist in three clusters of materials that are attached as Appendices.
preparing for e-scape leaflet Appendix 12
setting-up the e-scape system manual Appendix 9
familiarisation activity script Appendix 26

To give a flavour of the technical training day, we attach the agenda for that day as Appendix 25.



Snippets from the “user guides”

Since this was the last opportunity that we had to interact with the teachers in advance of the activities being run, we also created a pack of materials that they could take away with them and use for their preparations for the testing.

• a learner training guide so that teachers could familiarise learners with the hand-held device in the afternoon prior to the activity. See Appendix 26.
• a power-point presentation (to be used in the familiarisation session) that included images of all the interfaces that learners would have to operate. See Appendix 10
• a step-by-step overview of the whole activity – with indicative timings. See Appendix 13.
• the final form of the 'pill-organiser' learner workbook. See Appendix 15.
• the final form of the teacher administration script. See Appendix 14.

We believe that the training was effective – in the sense that the teachers have all subsequently successfully run the activities as required. However in order to understand better the impact of the training, we have received completed questionnaires from all of the teachers detailing their reactions to the training, to the activity, to their learners' performance, and to their attitude towards further e-scape-style activities in their schools. The data from these questionnaires are presented in section 8 below.

Before moving on from this matter of training, it is important to note that the teachers were NOT selected according to their familiarity or capability with ICT. We were primarily interested in their capability to run interesting design tasks with their learner groups. As a result of this priority it would be fair to characterise the group as very mixed ability, varying between – on one hand - complete ICT novices and – on the other – experience and expertise in normal d&t ICT systems, including e.g. CAD/CAM and presentation software. Whilst some were familiar with using VLEs, none of the teachers was familiar with using hand-held devices in the classroom, nor with using wi-fi systems to manage classroom activities. All this was new to all of the teachers.

the regional support team

One of the strategies that we adopted to support and monitor the schools was based on the regional groupings. In each of the regions we invited a support person to be the liaison and 1st point of support in the event that they needed it. The role that we envisaged is outlined in the briefing paper in Appendix 27 and is based on the idea that we wished to distance the research team from the operation as far as possible. As part of the *scalability* element of our brief, the research team would run the training, but thereafter we wanted the schools to be as self-sufficient as possible. The regional support person enabled us to distance ourselves from the schools without cutting them off from support.

The cold schools were organised into three regional clusters and each had a designated support person

- • South-West cluster (North Somerset/Bristol): David Perry
- • West-Midlands cluster (Shrop/Worcs/Birm): Paul Clewes
- • North-East cluster (Northumberland): John Chidgey
- • London / South-East cluster Tony Lawler

The support task:

• The support persons should attend the two training days provided by the central e-scape team;

• Beyond this, the support persons are the 1st port-of-call for teachers and schools seeking assistance during the preparations for the test activity. It is a *responsive* role, providing (as far as is possible and reasonable) whatever support they request.

• To enable us to document the support provided, support persons should keep a log of contacts with the schools, identifying the nature of the help provided and the time commitment involved.

• Support persons will be paid at the standard rate of £300 per day + reasonable expenses.

We hoped that the schools would need no further support – and that the support person would in fact be merely a conduit for organising dates and other administrative details. But this was something that had to be tested. The report in section 8 below is from Paul Clewes the W Midlands support member.

7. running the e-scape tests June/July 2008

the testing schedule for design & technology

The broad plan for testing in schools was that the e-scape system development would be completed by Easter 2008; that all the training sessions would be completed by half term in the summer term; and that the testing would take place in the 2nd half of the summer term, i.e. weeks of 2nd, 9th, 16th, 23rd, 30th June and 7th and 14th July.

The rationale for this timing has evolved over many years of working in schools, and specifically in relation to the availability of test-age groups (and their teachers). GCSE coursework assessment demands on teachers are at their peak in April / May, and in the 2nd half of the term we have consistently found it more possible to extract whole groups of GCSE age learners for extended projects (like our 2 morning, 6 hr task). We did not underestimate the difficulties of doing this however, and in the event the teachers reported a mixed bag of school cooperation. In the questionnaire to teachers, we asked specifically *"How easy was it to get learners off timetable for the activity?"* and their feedback was as follows:

very easy	*= 7*
easy	*= 3*
a bit difficult	*= 7*
very difficult	*= 1*

Whilst all sorts of competing activities sought to block it out (language oral tests / activities weeks etc), the key factor making it possible was the support of the head-teacher, the SMT and the department.

During the seven-week testing window, we planned to run the four sets of 'cold' school kits through 3 cycles, allowing us to operate in 12 schools. The three cycles were during the weeks of 9th June, 23rd June and 7th July, and they were centred in different geographical areas as follows:

9th June: West Midlands

- Droitwich Spa High School
- Walsall Academy
- Burton Borough School, Newport
- Abraham Darby School, Telford

23rd June: Northumberland

- Bedlington Community High School
- Coquet High School (Technology College), Amble
- Duchess's High School, Alnwick

7th July: South West

- Priory School, Weston Super Mare
- Clevedon Community School
- Gordano School, Portishead

During the 2nd and 3rd cycles (when the regional cluster only needed 3 sets) the 4th set of kit was available for us also to run tests in the two schools in the London area, Charles Darwin School in Biggin Hill, and Jo Richardson School in Dagenham. In all cases, these schools are mixed comprehensive schools.

science and geography testing

The brief that we gave to the science and geography development teams required the science and geography tests to be run in three schools at some point during the summer term. We left the teams to negotiate their own arrangements – assuming that they would have their own relationships with schools and teachers. And so it proved.

science schools

The three science schools were centred on Sheffield and Bath – the respective locations of the two science developers. The tests were run in the schools as follows:

- Birley Community College Sheffield – 23.6.08
- St Gregory's Catholic College, Bath – 8.7.08
- Ralph Allen School (specialist science college) 9.7.08 and 10.7.08

geography schools

A number of schools was used by the geography team during the activity-development process, and most were based around Bath and S Wales. The S Wales location was because of the particular cooperation between Fred Martin – the lead developer - at Bath Spa University and a teacher in Pencoed School Bridgend. The activity that became the focus of the geography test was based on the geographical demands of 're-branding' Porthcawl. Apart from the schools used for development purposes, the three test schools were all within easy travelling distance of Porthcawl (for the fieldwork element of the task) and these schools – and their test dates - were as follows:

- Dyffrn school: Margam, Port Talbot - 30.6.08
- Y Pant School: Pontyclun. - 8.7.08
- Llantwit Major School: Llantwit Major- 14.7.08

8. the response in schools

A variety of data types were collected during the running of the test activities in schools. At every session we had a member of the TERU team attending the activity. This was partly to reassure teachers that we could help with the technology if it should prove necessary (it rarely was), but mainly so that we could observe the process underway. For this purpose we designed an observation protocol recording the events of the activity at points throughout.

We also asked teachers – and all the learners – to complete a questionnaire at the end of the activity. The learner questionnaire was focussed on the handheld technology and on the extent to which they felt it had enabled them to do their best during the activity. The teacher questionnaire was more broad ranging, asking about the training; the set up of the room and the activity; the running of the activity; the technology; the response of their learner group; and their preparedness to do it in the context of real GCSE examinations.

Finally we asked our regional support team to send us reports on the activities that they observed – and their thoughts about it.

learner questionnaire

We have explained above (section 4) that two handheld devices were used in the tests; the HTC 'advantage', sometimes named 'ameo', and the Toshiba 'portege'. They have some differences in design and some different features, but the overriding difference is in the screen size; the ameo being significantly larger. The questionnaire data from learners is influenced by this difference and has therefore been analysed separately.

using the Ameo

212 students (130 male, 80 female) completed the Ameo questionnaire and the vast majority were very supportive of its value as a designing tool.

- 97% agreed or strongly agreed that the device was easy to learn to use
- 87% that it was easy to use to develop ideas
- 97% that it was good for photographing emerging work
- 92% that it was overall a good tool for designing
- 95% that it was fun to use

In free-response mode, the functions, or aspects of performance that learners particularly appreciated are listed below. The fact that it was fun/easy to use was mentioned 147 times (94 male / 53 female). All nine items were mentioned more than 20 times.

Thumbs up	M	F	ALL
fun / easy to use	94	53	147
good camera	44	23	67
small & compact / stylish / portable	27	12	39
drawing / paint tool	23	15	38
new / different	13	20	33
not just for writing on / interactive / can do a range	15	14	29
voice recordings	17	8	25
video	13	12	25
transcriber / text tool / keyboard / writing	9	15	24

The list of functions or aspects of performance that learners particularly disliked are again listed here. Only 7 items were mentioned more than 20 times and these were dominantly about technical matters (crashing / loading / deleting) that have been ironed out in subsequent iterations of the software. As an example, many 'freezes' were attributable to learners opening other programmes (eg games that it proved impossible to remove) in the device and overloading the available memory. The gender difference in these data is interesting in this regard. We generally overcame this freezing problem by explaining it openly to learners and requesting their co-operation. Nevertheless, the issue of technical problems giving rise to missing data was important enough to justify specific analysis, and this is included in section 12.

Thumbs down	M	F	ALL
sometimes crashed / froze	60	25	85
hard to type / write	29	17	46
timings / time bar	22	23	45
slow to load	25	16	41
hard to draw	18	14	32
Things can deleted / not save / no collaboration	16	13	29
gets some time to get used to / complicated	13	14	27

using the Toshiba

41 students (20 male, 21 female) completed the Toshiba questionnaire and again the vast majority were very supportive of its value as a designing tool.

- 98% agreed or strongly agreed that the device was easy to learn to use
- 98% that it was good for photographing emerging work
- 95% that it was a good text tool for explaining ideas
- 95% that it was fun to use

In free-response mode, the functions or aspects of performance that learners particularly appreciated are listed below. All 8 items were mentioned at least 5 times.

Thumbs up	M	F	ALL
camera / picture tool	12	12	24
fun/easy to use	12	8	20
good experience / interesting / different	3	7	10
video	5	3	8
voice recording	4	3	7
better than doing things on paper / not writing loads	2	3	5
developing ideas	2	3	5
touch screen		5	5

The list of functions or aspects of performance that learners particularly disliked are again listed here. Eight items were mentioned more than 5 times and these were typically either about technical matters (crashing / slow) or were about the smallness / fiddlyness of the device

Thumbs down	M	F	ALL
slow / unreliable	11	8	19
difficult to draw	6	7	13
fiddly / tricky / hard / hard to understand	3	8	11
mixed up comments / partner work	3	7	10
crashed / froze	3	5	8
keyboard a bit complicated	3	3	6
voice recording	2	4	6
could have done better / got things wrong on it	2	4	6

comparing the Ameo and the Toshiba

It is interesting that the Toshiba gets significantly less support than the Ameo for some of the key design functions:

Developing ideas:	Ameo 87%	Toshiba 71%
Sketching:	Ameo 65%	Toshiba 37%

Overall the Toshiba is seen as a good tool for designing by only 76% (as against 92% with the Ameo). This seems to be a matter of size and manageability. Having said this however, it is reassuring that the vast majority of responses from learners have been positive about the use of the handheld devices as a means for enabling them to construct their design portfolios.

teacher questionnaire

The teacher questionnaire was completed by 18 teachers (14 male 4 female), and of this group, 17 / 18 had 'limited' or 'none' experience of using handheld devices in the classroom.

Concerning the training they received from the TERU team:
17 / 18 were 'confident' or 'very confident' about preparing handling collections.
17 / 18 were 'confident' or 'very confident' about preparing modelling kits
17 / 18 were 'confident' or 'very confident' about running the activity
11 / 18 were 'confident' or 'very confident' about handling the laptop/handhelds..

And in the event …
16 / 18 reported it was 'easy' or 'very easy' to set up the room for the activity
15 / 18 reported it was 'easy' or 'very easy' to manage the activity via the laptop
18 / 18 reported it was 'easy' or 'very easy' to cope with the handhelds

Concerning learners performance …
16 / 18 reported that learners did 'better' or 'lots better' than expected, noting particularly the *pacy timings* and the *variety of ways* for learners to respond

17 / 18 reported that learners motivation was 'better' or 'lots better' than normal, noting particularly the clear targets and the range of activities

For the future …
17 / 18 would 'probably' or 'definitely' do it again, noting particularly that it was innovative and that students enjoyed it and stayed on task throughout

The following statements are direct quotes from the free-response invitation at the end of the questionnaire. We have included them here in their entirety since they include not only some very honest appraisals of the system, but also human reactions to the task, to the kit and to the general demands of the research setting into which they were thrust.

Overall I have enjoyed being part of the research and I believe the students gained from taking part and they may be able to make their future D&T experiences more varied and richer for the experience. Traditional values / attributes are challenged in this mode of working. Assessment process looks interesting and I would be interested in following this though.

Absolutely loved it. Worked very well. Shifted class around from traditional skills to a better glimpse of design capability …. Raving about it - any help in the future ....

Some groups would struggle with the amount of responsibility they would have - therefore, more training / experience would be needed by them. Would we have the same control over a student using their own mobile phone? Lots of extra stuff to 'control'. Training them to use responsibly.

Thoroughly enjoyable for both students and staff. Students remained focused for the full session and produced work that is to the best of their ability.

An inspiring and worthwhile exercise. Students see the value, are motivated and enjoy using it.

Troubleshooter guide i.e. "what if …" Ist Day - Nerves, how will students respond, will the technology work, reading from a script? How do you motivate? Emphasis the pressure is not about marking but about being creative. After initial training day it seem that tech would confuse some and small glitches could throw timing and flow of the activity. No need to worry - there were occasional problems, things not being saved correctly, peer comments not recorded etc. These were feasible to solve. PDAs quickly became a device or tool for recording their work, The novelty had worn off but you sensed that they enjoyed working in this way. The short timings propelled the task forward and put real pressure on students to perform. Most did but a few complained that there was not enough time to complete their tasks. however if extra time was allotted how much more would have been done? Not in proportion. The device with its timed screens created episode deadlines with review points. Timings had to be changed to accommodate our short morning. Unfortunately this took time off modelling ideas. Most sketched initially but then mainly modelled. Interest from school - HT, Deputy Head, head of IT …

I thoroughly enjoyed planning and conducting the trial - as did the students. The whole process from training to implementation was well organised, well structured and well carried out. Students were very enthusiastic about taking part in the trial despite the conflicting demands put upon them with orals and Y10 internal exams looming next week. They all seemed to enjoy the experience and produced some very creative and well considered ideas. Interesting that one or two of our lower achievers were the ones who were most at ease and competent with the devices. I was welcoming of the support at the start of the day 1 but thanks to the reliability of the system this wasn't needed much once the trial was underway. Many thanks for considering us and using us for the trial and for the kind words which were said about our students. Looking forward to experimenting with a similar approach ourselves and the 'pairs marking' in September.

I have thoroughly enjoyed using the e-scape system and it has obviously become easier to set up and run activities with practice. What a powerful tool for a teacher! I particularly like the way sound files, photos and videos can be simply integrated into a normal classroom situation, making it possible to access thoughts and ideas of every individual (including SEN) rather than just those who it is possible to get round and talk to - really powerful for assessment for learning. Tremendous potential for all curriculum areas and age groups, including cross-curricular work.

report from the W Midlands regional co-ordinator

Background

My contact with the e-scape project has been on-going over a number of years. In my previous role as a county D&T advisor I had recommended local schools to the TERU team and I had visited the phase 2 trials and the report back day. I was delighted therefore to be asked to suggest further schools for phase 3 of the project and to take a more supportive role. The schools that I suggested included a new and very over-subscribed school with a high levels of pupil motivation, as well as a school which had recently emerged from 'special measures' but retaining a high percentage of challenging pupils. The other schools might be regarded as being somewhere between and would have features which might be seen in many schools across the country. All of the schools suggested have staff who would be willing and capable of meeting the demands of the project and have 'good' D&T departments. Following an initial invitation to take part, none of which were rejected, all four schools were visited by the e-scape team. In some schools either another teacher or a member of SMT also joined these introductory meetings. Before the week of the trials the teachers and I then met with members of the e-scape team for a regional training session in the West Midlands as well as a national session in London, elements of a team spirit were emerging. One of the training days was focussed on the preparations that a school might make in terms of the pupil groups, timings, modelling resources, handling collections. The second training day specifically concerned the use of the portable computers and the wireless connections for the trials. My role was one of providing local support to these 4 schools. The offer of this support was made both in person as well in the e-mail communications that were established.

During the week of the trial I visited three out of the four schools in the West Midlands region. These events were recorded photographically and the opportunity was taken to discuss elements of the project with teachers, classroom helpers, e-scape observers and pupils taking part in the trials.

Observations:

From a personal view I think that the principle of having focussed innovation challenges as a regular part of pupils D&T experience is one that all schools should consider. In my view too many schools offer an 'engineered success' D&T experience to their students and rarely offer anything in the way of opportunities for them to show innovation or genuine creativity with an element of risk, where real success or failure can happen. The advantages of a fixed

timeframe innovation challenge include; the greater pace that all pupils work at, the 'walled garden' opportunity to design and make within constraints of the available time, resources, materials, and a lack of a long term investment in a project. This gives the pupils a greater freedom to take risk and to experiment with their designing and making.

My observations of the latest phase of the e-scape project challenges in three of the West Midlands schools were that:
Staff and pupils acquired the necessary skills to work the ICT very quickly and with no significant problems. The wireless system used appeared to be reliable and no significant breakdowns occurred. Many teachers were able to provide support to pupils and to solve the issues which occasionally arose. Most difficulties appeared to be caused by pupils over enthusiastic closing and saving programmes in a different way than was required. The actual use of the hand held computers added a significant dimension to the event in that it was 'special'; the computers were so portable and versatile that they were used for different purposes over the sessions which maintained the interest level. Pupils quickly overcame the embarrassment factor of talking to or videoing with the computers as the whole class were doing the same task at the same time. Pupils took good care of the computers which were kept on the same tables as serious modelling was taking place with such things as knives, scissors and glue. Significantly, no damage was suffered to the computers. They were the recording device which became the 'virtual exam paper' and so gained a respected item status because they were the facility through which pupils communicated their progress. Teachers also found the approach of facilitator (but not advisor) a different role from usual. This was, I suspect, a refreshing change from the more interventional, pro-active and supporting role that they might normally adopt. To a certain extent the pressure is off them as teachers and on to the pupils to show what they could do independently. The teachers interpreted their 'scripts' generally very well and provided the resources necessary for the activities to take place smoothly. The handling collection appeared to have been more influential in pupils designing when it was a smaller collection replicated on each group's table rather than a larger collection on a side table.

Summary:
The schools were enthusiastic about taking part in this project, trying out (in a D&T context) things which were innovative and focused on encouraging pupils themselves to be innovative. Whether this enthusiasm by teachers and pupils would be maintained if the process was repeated (with different design challenges) would be interesting to see. One hopeful outcome would be that pupils simply got better at being innovative and being prepared to be challenged. Some schools that I am aware of (BSF) are reviewing their whole school ICT provision and are considering a more widespread adoption of this technology in many curriculum areas.
I observed a well-managed project which was delivered equally well in schools with differing circumstances. I consider the outcome of these trials was that pupils' creativity was given an opportunity to flourish within a limited timeframe and that all pupils were better motivated, more involved and produced more original ideas than in any comparable designing situation.

9. judging the web-portfolios

recruiting and training judges

Our basic plan for the recruitment of judges was informed by the 'scalability' element of the brief for phase 3. If the system is to be used as part of a national examination system, then the judging process needs to be carefully accounted in terms of time spent in relation to existing systems.

Currently, when teachers enter learners for d&t coursework assessment, they are automatically committed to making assessments of learners' coursework portfolios in their own time. We know that teachers allocate large chunks of time to this task, and this is entirely additional to any examination marking that they might take on. Accordingly, we worked on the idea that for every group of learners in the e-scape pilot of 2008, we would recruit their teacher as a 'pairs' judge. Since we had 18 schools represented in the pilot that gave us a good start in putting together a team of judges. Additionally we found ourselves with an embarrassment of riches as more and more people volunteered to take part for the experience of seeing it in action. In the end we

had a team of 28 judges, including two from Ireland, one from Israel, three independent d&t consultants (former advisers), the central research team of six, and 16 teachers from e-scape pilot schools.

Our plan was to train this group in a one-day session in central London. The training day was divided into two broad sections:

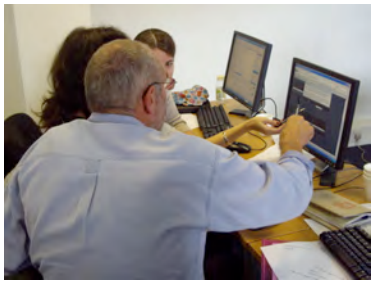
• concerning the ***evidence*** in portfolios – and making ***holistic*** judgements

• concerning the ***pairs engine*** and specifically the judging interface

The training day took place in an IT training suite. This had an unfortunate – rows-of-desks – layout that we were unable to alter, but apart from that shortcoming it proved satisfactory. Essentially the morning was spent scrutinising the evidence initially in paper-based portfolios in relation to the key criteria of performance. Teachers worked in pairs to discuss and hopefully agree on the quality of evidence in particular portfolios. Having got to grips with the criteria and the kinds of evidence that they were likely to have to deal with in the portfolios, we spent time debating and practising the exercise of holistic judgement. Thereafter – in the afternoon – we introduced the pairs engine and the interface that they would be required to manage.

It did not take long for judges to become familiar with the engine interface, since it operates at a fairly intuitive level. However the ***digital portfolios*** offered considerably more evidence than their paper equivalents – not least through the audio files that recur throughout the working session and through the video presentation by learners at the end of the development process.

Once again the judges were paired and spent an hour scrutinising the evidence in the portfolios and rehearsing the process of holistic judging. When they expressed themselves comfortable with the process – we asked them to work independently and tackle a series of pairs judgements.



We had once again pre-selected the 20 portfolios from the initial trial and with 28 judges it only took an hour for all the judges to complete their allocation of paired judgements. At the end of the process we had generated (again) the rank order that we had already agreed in the TERU office trial, and we were able to show this to the judges and share their thoughts about the appropriateness of the rank order.

A month later, at the end of the judging process (having completed their 127 paired judgements) we asked all judges to complete a questionnaire about it that included a section on the training. The following comments are representative of their responses to a question about whether the training had been effective.

Yes (RM)
yes – plenty sufficient and effective; instructions documents very helpful too (RW)
Yes! The transition from criteria based marking to holistic judgements wasn't too much of a challenge for me. I think that much of this can be put down to the initial training in making the pairs judgements (DW)
The training day in London was thorough and gave me confidence to tackle the judging. I felt that all aspects from the technical to the criteria for assessment had been clearly dealt with. (AM)
Yes. Also the system was very user-friendly (PH)
Yes certainly (TL)
*Yes, but it still took me some time to overcome ingrained, detailed examination of the folios and being prepared to adopt the holistic judging system. This became much easier a) after having completed a number of comparisons, and b) once the chain started. (DH)**

** this reference to the 'chain' starting is explained below*

At the end of the training day we issued all the final rules for the conduct of the pairs judging process itself – including logging in and passwords to maintain absolute security in the system. We needed one further week to finalise the appearance of the portfolios in the engine, and we were therefore ready to start the process on Monday 6th October. We allocated a three-week window for everyone to complete the 127 judgements that were required. This figure arose from the following simple arithmetic:

- the number of portfolios = 350
- the number of comparisons for each portfolio = 20

= 7000 portfolios to be seen

= 3500 when seen as pairs to be judged

Hence, with a team of 28 judges

= 127 paired judgements for each judge

judges' reflections on using the system

At the end of the judging process, we asked judges to give us some feedback on their experience of the process – and their views about it. We used a 'free response' questionnaire structured around some broad-ranging questions.

Did the e-scape portfolios enable judges to 'see' learners' capability? The overwhelming response of the judges was **that learners' capability was very evident** in the portfolios.

> *The variety of media used within the portfolio seems to give everyone a fair chance to display their design thinking regardless of their ability. (DW)*
> *Most certainly and in ways which previously we could only dream about. (TL)*
> *Overall yes. There was plenty of scope for the students to show a range of capabilities from generating ideas to evaluating the ideas of their peers, from drawing skills to modelling skills. (AM)*

And as a result, the consensus of opinion amongst the judges was that the 350 portfolios represented a **broad spectrum** of capability.

> *It was fairly easy to see that there was a range of performance evident in the portfolios. (DW)*
> *Yes, from little creativity and growth to a very good level of both. (DH)*
> *Oh yes. Quality did shine through, as did mediocrity (TL)*

I could see progress / or when students had simply abandoned one idea and stared another. It was also evident whether ideas had 'grown' and what inspired the students. (HW)
A wide range of ability and completion. (RM)
Yes, without doubt. Surprisingly pupils' attitudes were also often evident. (PH)

The main **strengths** of the e-scape e-portfolios were characterised in three ways. *First* in terms of presenting the 'big-picture' of learner performance; *second* in terms of capturing the critical ephemeral evidence alongside the more obvious artefact evidence; and *third* in the ability to see the growth of ideas through the activity.

Portfolios displayed in this way have a huge advantage in that "the big picture" can be seen immediately. It's very easy to get a "feel for the project" that would not be possible unless the whole project was displayed on one single screen. The ability to "dip in and out" of the different sections enabled me to reinforce my holistic mental picture of the project. (DW)
The ability to hoover up ephemeral evidence of designing and creativity (TL)
Improvement of ideas (VG)
Growth and development of the design ideas became quite easy to visualise … (DW)

Two areas of **weakness** were noted in the portfolios: *first* technological issues about the functioning of the portfolio system itself, and *second* concerning the incompleteness of some of the learners' portfolios.

The way that the text was enlarged when hovering over it with the mouse was effective, it is a pity that the pictures had to be opened in a new window. (DW)
Learning to operate it on screen, but that was only in the beginning (TL)
Incomplete portfolios … (RM)
Some students do not do themselves justice because they have not recorded the evidence of their creativity and progression (DH)
Some candidates undermined their position by not submitting key elements such as the video (DP)

When comparing **e-scape judging** to more **conventional GCSE marking**, judges were unequivocal about the ease of using the judging system.

Easier assessment; no need to calculate grades and points (RM)
Speed of judging (VG)
It's brilliant not having to carry around boxes full of portfolios/folders and practical work! Speed - all the paperwork, calculations, moderation etc. are not needed. (DW)
Quicker to make assessments (RM)
much, much faster … less scary (re individual marker impact on individual learner life chances)… get a whole view much more readily (RW)

Moreover, many commented that the system was **more fair** since judges are not bound into the details of exam board criteria – but can (through the holistic judgement) reward qualities that are important but not pre-specified.

Not having to deal with "criteria" that often don't reflect the full range of design and technology activity. (DW)
GCSE marking relies heavily on a tick box assessment of a pupil's work. It can be frustrating when confronted with an excellent piece of designing and making that does not meet the exam board's criteria. Too often the linear pattern of coursework requires the assessor to jump back and forth to find the marks that a student deserves. The e-scape judging is so simple in comparison. (AM)

Yet another source of **enhanced fairness** was seen to result from the accumulation of judgements from multiple judges

The judging system feels to be fair; it doesn't rely on only one person assessing a single piece of work. It removes virtually all risk of bias.... It feels safe knowing that even if you make a mistake in one judgement it won't significantly make a difference to the outcome or grade awarded to the student as other judges will also assess the same project. Also knowing that the system automatically checks the consistency of the assessor's judgements again reinforces the feeling of fairness that this process brings. (DW)
less scary (re individual marker impact on individual learner life chances) (RW)

We were very aware that a critical difference between e-scape judging and GCSE marking is the centrality of the **holistic judgement**. We were pleased to hear that judges felt that it was easy to hold a sense of holistic capability and use it to reward good performance.

feels like might get judgement more 'right' than e.g. by starting from an accumulation of sub-judgements… feels right thing to do – that's what D&T is about (RW)
Making holistic judgements meant that I was not forced to give credit to an apparently well-designed project that was completely unrealistic in terms of being an actual product. (VG)
It values and consolidates the beliefs in what is 'goodness' in Design and Technology..(TL)
It gives more appropriate results than atomised approaches which can lead to inaccurate overall assessment especially when the overall attainment is more than the sum of the parts. This often happens when the various elements of a designing process come together in a successful outcome that outstrips the quality of work in any (or all) the parts of the process. (DP)
One of the major strengths of holistic judgements I see is its flexibility… in which you can give credit to students for what they have actually done rather than whether they are able to "tick the boxes" to match a set of assessment criteria. (DW)

There are some caveats however about holistic e-scape judging, and one that emerged concerned the lack of **formative feedback** from the system to the learner.

it does not really produce any evidence or feedback that could be given to learners/candidates or schools/centres at the end of the process. This lack of feedback would be particularly difficult for a student or school that is struggling to "make the grade", they would have few, if any, indicators of ways to make improvement. (DW)

In fact there are many solutions to this matter – and the problem does not need to exist. Whist the current e-scape judging system has been optimised for summative assessment purposes, the next phase of this work explores exactly the formative benefit of judging for (and by) learners. In fact a whole new research strand is being conducted (in Scotland with LTS and SQA) to explore the classroom benefits for learners and teachers.

It was with this in mind that we asked judges about the *potential* classroom benefits of using the e-scape judging system as a natural part of classroom / studio practice. The responses were overwhelmingly positive.

would support learning – e.g.:
• would encourage learners understanding of capability
• would encourage learning through critique – and hopefully expose to what 'yawn' and 'wow' designing looks like at different 'maturity levels
• would be cohesive with [my view of] D&T philosophy – e.g. learning is situated, socially constructed

• could support policy efforts to involve learners more in shaping curriculum (in its widest sense)
• could support e.g. PLTs evidence (RW)

Absolutely! Student ownership of the assessment process; learning in the hands of students; students setting their own assessment/success criteria; empowerment. Teacher satisfaction; greater focus on ***learning*** *rather than teaching and assessment. (RM)*

Yes please! Any system that takes away the drudgery of continually marking work and improves the quality of assessment has to be a priority for any teacher! If students can take an active part in the judgment processes then a whole new set of learning tools come into play, such as self and peer assessment, being able to see examples of work from other students … with further development and refinement, could bring a whole new perspective on assessment, marking, recording, and reporting (DW)

I think pupils having an e-portfolio such as this to look at would certainly support their learning and help them improve their work and develop it further. (DH)

And .. a few final thoughts on the whole process

Overall it was fascinating. I think there is huge potential for using e-portfolios in many curriculum areas providing opportunities for a combination of self-assessment and teacher formative assessment – audio and videos provide opportunities for accessing each student's thinking, almost equivalent to talking to each individual. (PH)

This is a visionary use of modern technology to support what good teachers of design and technology have always believed in but have had to stretch the examination systems to show. With this, good Design and Technology can be rewarded as a result of (and not in spite of) the examinations (TL)

the learner voice emerges (RW)

I genuinely think that students would enjoy the freedom that this gives to their own creative expression. (AM)

Over time, this assessment technique would be likely to enhance the quality of practice of D&T teachers. Where current assessment techniques encourage a reductionist approach exemplified by algorithmic approaches to designing and a lot of pretty-but-shallow work, this technique should encourage learner collaboration; stronger independent working; more reflective ability and self-evaluation; and the ability of students to discriminate good design work from poor work. For these reasons in particular I consider that this approach has the potential to be a much improved method of assessing large numbers of students when compared with existing methods. (DP)

10. analysing the judgement data

Throughout the e-scape project, and particularly in phases 2 and 3 we have been working in close association with Alastair Pollitt. He first introduced us to Thurstone's 'comparative pairs' judgement and in 2006-7 he organised the statistical methodology for running the judgement process (through 3 stages) and subsequently analysed the emerging data.

In this phase 3 of e-scape, Pollitt has been closely involved in the development of the 'pairs engine' that now effectively (and dynamically) manages the judgement process and he has also completed the analyses on the resulting data. Rather than pick and choose from his report to us on the outcome of his analyses, we felt it more appropriate to include his report in full.

report from Alastair Pollitt on the d&t judging

Report on the analysis of the e-scape trial. Alastair Pollitt.

Collecting and analysing the judgment data

Preliminary

In the previous trial (*e-scape2*), data were collected and analysed in three distinct phases. First, eight judgements were collected for each portfolio with the pairings made entirely randomly. The data were then analysed to create a preliminary scale, and this scale was used to choose more appropriate pairings for the second phase. After a further eight 'targeted' comparative judgements had been collected for each portfolio the accumulated data were analysed again to give a new set of value parameters. A smaller final phase was used to collect data for specific purposes, and a third analysis produced the final values for the parameters.

In this trial, of a pilot operational system, judgement data were analysed more frequently, and in just two phases. First, to get the system started a "Swiss tournament" system was used. In the first round 176 random pairs were judged, and the portfolios put into two sets – 'winners' and 'losers'. In the second round 'winners' were compared only to other 'winners' and 'losers' only to other 'losers'; this made three sets of portfolios with 2 or 1 or 0 wins. After each round of 176 judgements the portfolios were again sorted in this way, so that after six rounds there were seven groups.

These groupings generated initial estimates for the full estimation process, which was introduced after the seventh round of comparisons. From this point on, after every 176 new judgements added to the data set there was *on average* one new judgement for each portfolio, and a new calibration was generated. In all, there were about 17 and a half rounds before the process was stopped.

Evaluation: Initial phase

The efficiency of the Swiss system can be compared to that of a random system, as used in the previous trial.

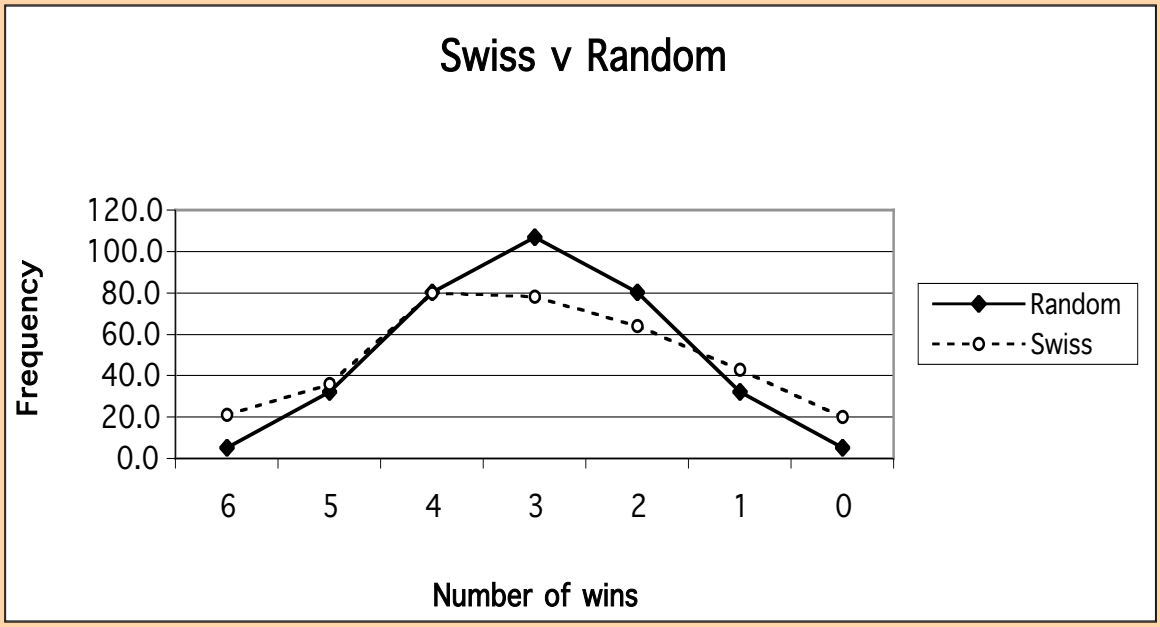


Figure 1: Score distributions for Random and Swiss systems

If pairings are always random then the prior probability that any portfolio will win any comparison will always be 0.5, and a simple binomial calculation predicts the distribution of scores 0-6 as shown by the "Random" line in Figure 1.

The actual distribution is shown by the "Swiss" line. There were more portfolios in the extreme categories than Random would have given, and therefore there is more discrimination in the system to start the proper estimation procedure. Correspondingly, the standard deviation of 'scores' has increased from 1.22 in the Random data to 1.55 in the Swiss.

We can also estimate directly the "efficiency" of the pairings that were made by the Swiss system. In statistical theory, the *information* provided by a comparison is given by $p*(1-p)$, where p is the probability that the first portfolio will be judged better. The comparison is maximally efficient (100%) if the *true* values of the two portfolios' parameters are exactly equal, since p and $(1-p)$ will then both equal 0.5. Any value of p that is more or less than 0.5 will result in a smaller amount of information[1]. When the pairing is chosen, of course, we do not know the true values and can only use the latest estimates.

From the analyses of the data at different stages of the trials, the average efficiencies[2] for each comparison in different trial parts were estimated as follows:

escape2:	8 Random untargeted comparisons:	34.3%
escape2:	8 Fully targeted[3] comparisons:	79.0%
escape3:	6 Swiss comparisons:	50.8%

The efficiencies in the targeted comparisons are, of course, much more efficient than the random ones, but the figures also show that the Swiss approach is considerably more efficient than a random approach to setting up pairings – and, unlike true targeting, it can be used before any proper estimation has been attempted. In effect, the Swiss tournament system achieves an approximate targeting at almost no computational cost.

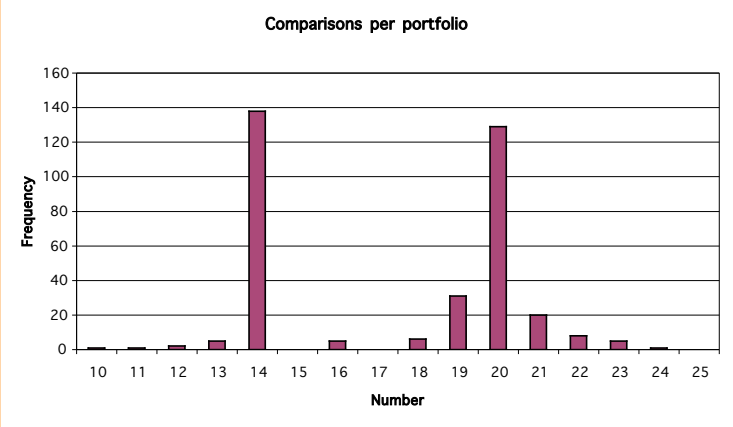
Evaluation: Main phase

Throughout the data collection period, the targeting of comparisons steadily improves. In the very first round pairing is purely random; in the second (Swiss) round the pairings are drawn from within the two score categories; in the third they are from the three categories, and so on up to six rounds and seven categories. After this point, all of the pairings were targeted at the 89% efficiency mentioned earlier, but the targeting continues to improve as accumulating more data makes the estimates more accurate.

It is a feature of Thurstone's method that every paired comparison is independent of the others, and the data from both phases are simply combined for analysis. The final data set therefore consisted of 3067 judgements, made by 28 judges, and involving 352 portfolios.

Numbers of comparisons per portfolio

Figure 2 shows how many judgements were made per portfolio.



[1] For reasons discussed in the report on the previous trial, we chose to aim for comparisons that were not quite theoretically optimal – at about 89%.

[2] These are calculated using the final parameter estimates.

[3] These comparisons were targeted as precisely as possible using the parameters estimated from the first eight random rounds.

Figure 2: Distribution of comparative judgements

On average, portfolios were judged 17.4 times, but most were seen either 14 or about 20 times. There seemed little difference between those seen only 14 times and the rest except, of course in the average size of their estimation standard errors.

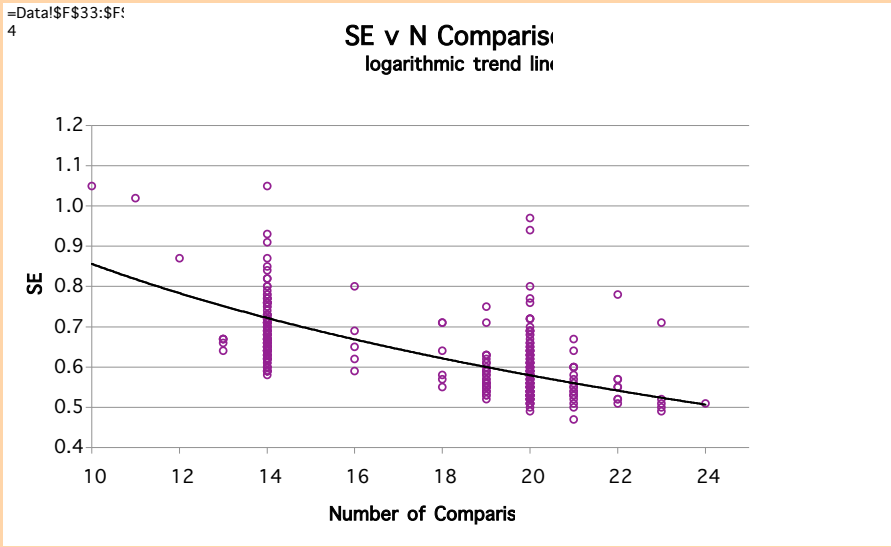


Figure 3: Standard errors of estimation for different numbers of judgements

Estimation

The analysis begins by constructing a square matrix showing how often every possible pairing of portfolios was seen by any judge. In this case there were 352*351 = 123,552 possible outcomes, or entries in the matrix. Each comparison makes two entries in the matrix, one for whichever portfolio 'won' and another for whichever 'lost'. But since only 3,069 comparisons were made, at least 95% of the matrix cells must remain empty, even if the other 5% contained only 1s and 0s. There was some concern that the iterative procedure that estimates the parameters would collapse when based on such a sparse matrix.

In fact, it did not. The history of the estimation process is shown in Table 1:

Table 1: *History of the estimation process*

Summary of Pair Iteration Process: Parameter Estimate Shifts

Iter	RMS	Max	Object	Var
1	0.7670	2.3617	p202	0.5869
2	0.4591	1.4092	as20	1.3799
3	0.3303	1.1324	p202	2.2109
4	0.2596	1.0410	p202	3.0134
5	0.2105	0.8937	p202	3.7600
6	0.1708	0.6550	as21	4.4288
7	0.1403	0.5416	as21	5.0148
8	0.1171	0.4383	as21	5.5316
9	0.0982	0.3477	as21	5.9852
10	0.0826	0.2723	as21	6.3808
11	0.0697	0.2121	as21	6.7245
12	0.0590	0.1820	cd07	7.0222
13	0.0501	0.1586	cd07	7.2796
14	0.0427	0.1383	cd07	7.5019
15	0.0365	0.1206	cd07	7.6939
16	0.0312	0.1052	cd07	7.8595
17	0.0268	0.0917	cd07	8.0024
18	0.0230	0.0800	cd07	8.1257
19	0.0198	0.0697	cd07	8.2320
20	0.0170	0.0608	cd07	8.3238

Iter	RMS	Max	Object	Var
21	0.0147	0.0530	cd07	8.4029
22	0.0127	0.0462	cd07	8.4712
23	0.0109	0.0403	cd07	8.5300
24	0.0094	0.0351	cd07	8.5808
25	0.0082	0.0306	cd07	8.6246
26	0.0071	0.0267	cd07	8.6624
27	0.0061	0.0233	cd07	8.6951
28	0.0053	0.0203	cd07	8.7232
29	0.0046	0.0177	cd07	8.7475
30	0.0040	0.0154	cd07	8.7684
31	0.0035	0.0135	cd07	8.7865
32	0.0030	0.0118	cd07	8.8021
33	0.0027	0.0103	cd07	8.8155
34	0.0023	0.0090	cd07	8.8271
35	0.0021	0.0079	cd07	8.8372
36	0.0018	0.0070	cd07	8.8458
37	0.0016	0.0061	cd07	8.8533
38	0.0015	0.0054	cd07	8.8597
39	0.0013	0.0048	cd07	8.8653
40	0.0012	0.0042	cd07	8.8701

The key values, under 'RMS' and 'Max', show the average and the largest changes made to the parameters in each cycle. Sufficient stability was reached after 24 cycles, when these values were below 0.01 and 0.05 respectively. The report also shows that most of the difficulty in achieving satisfactory stability was caused by a single portfolio – cd07.

The outcome is a table showing the final estimate of the parameter value for each portfolio, together with its estimation standard error:

Estimates of object parameters

Number	Parameter	S.Error	Object Name
1	-2.160	0.540	ad13
2	-3.194	0.589	ad09
3	2.807	0.594	ad14
4	0.225	0.617	ad17
5	-1.767	0.578	ad02
6	-3.321	0.586	ad11
7	-1.567	0.589	ad05

etc

These values can then be sorted to give a final rank order of the relative quality of the portfolios, as judged by, in each case, about 117 judges. The full sorted table is included later in this report.

Quality control

Because every single judgement made can be compared to the outcome predicted (with the benefit of hindsight) from the final rank ordering, very detailed monitoring is possible of the consistency of the judgements made by each judge, and of each portfolio.

Table 2: *Summary of fit statistics for judges*

Number	Count	MnResid	UWMnSq	WMnSq
1	127	0.23	0.93	0.87
2	126	0.24	1.62	1.32
3	126	0.24	**6.82**	1.48
4	126	0.28	2.08	1.48
5	96	0.27	0.70	0.81
6	126	0.33	1.28	0.92
7	126	0.31	0.73	0.77
8	63	0.39	0.80	0.81
9	126	0.33	0.58	0.65
10	125	0.34	1.22	1.03
11	121	0.23	1.83	1.41
12	21	0.35	0.70	0.79
13	127	0.22	0.75	0.77
14	59	0.26	**8.55**	**1.96**
15	90	0.38	0.76	0.77
16	30	0.31	0.61	0.76
17	126	0.32	0.55	0.62

18	127	0.25	0.89	0.94
19	120	0.36	1.04	0.94
20	123	0.33	0.62	0.69
21	126	0.29	0.92	0.91
22	127	0.25	1.16	1.23
23	127	0.23	**5.24**	**1.60**
24	126	0.28	0.95	1.13
25	127	0.29	0.77	0.97
26	65	0.39	0.75	0.81
27	127	0.31	0.69	0.83
28	126	0.34	0.74	0.73
Mean:		0.30	1.58	1.00
S.D.:		0.05	1.92	0.32

The most important statistic indicating misfit is the *weighted mean square* (WmnSq). Theory predicts that this statistic should average 1.00, and in these data it does exactly that. The standard deviation is more difficult to predict: the usual way to interpret a mean square is as follows. Calculate 'Mean + 2xSD'; then any value greater than that criterion is considered to be showing a significant amount of misfit.

Here the calculation gives 1.64 as a criterion, and only one judge exceeds this – Number 14. It may be significant that this judge made only 59 judgements, while the others averaged 111.5, or almost twice as many – this misfit may be a consequence of inadequate experience. A second judge – Number 23 – comes close, with a mean square of 1.64. The *unweighted mean square* (UWMnSq) is a similar statistic, but is less stable and particularly sensitive to occasional extreme 'misjudgements'. Three judges show high values on this statistic. There is no surprise that the same two as before give high misfit values, but Number 3 also appears. Here the weighted statistic is acceptable, but the unweighted one is too high; this usually indicates just one or two highly surprising judgements – or an error in the data collection.
Overall the amount of misfit seems quite acceptable.

Evidence for fit or misfit can be summarised for portfolios as well as for judges. The summary of misfit for them is given below:

Summary of Fit Statistics for Objects

Numb Name	Parameter	SE	UWMnSq	WMnSq
Mean:	0.00	0.64	1.50	0.94
S.D.:		2.97	0.19	2.90	0.29

The weighted mean square this time averages 0.94, just a little lower than the theoretical expectation. The criterion for clear misfit would be 0.94 + 2x0.29, or 1.52.

Table 3: *Summary of fit statistics for portfolios (top part only)*

Numb	Name	Parameter	SE	UWMnSq	WmnSq
153	cq16	-1.40	0.69	2.19	1.78
155	cq06	5.07	0.87	4.46	1.74
186	dr14	-2.21	0.57	**15.69**	1.70
24	ap02	1.10	0.68	4.54	1.68
84	bb08	0.66	0.63	4.78	1.66
69	bd04	1.07	0.64	5.29	1.62
161	cq19	5.95	0.72	1.28	1.61
222	gd03	5.62	0.76	2.94	1.61
176	dr10	0.80	0.71	1.40	1.60
228	gd16	2.00	0.71	5.67	1.57
241	jr08	-0.36	0.68	1.94	1.57
338	ws03	1.90	0.67	5.51	1.57
98	bb01	0.93	0.68	1.62	1.54
197	ds12	-2.85	0.63	**27.11**	1.54
256	nc21	-1.04	0.64	2.29	1.54
70	bd02	0.09	0.72	1.73	1.52
260	nc11	-5.16	0.75	5.11	1.51
292	sk17	1.67	0.59	**12.62**	1.51
214	gd11	3.26	0.75	**17.83**	1.48
etc					

16 portfolios exceeded this level, or 4.5%, which is satisfactory for a 5% significance test. In an operational system, nevertheless, these 16 should be reconsidered to see if there is some significant reason for their misfit. Ten portfolios showed extremely high values of the unweighted mean square – four of them can be seen in this excerpt; these should be checked (if possible) to confirm that no error was made in data collection.

Overall summary

Table 4 estimates the overall quality of the process, in terms of consistency.

Table 4: *Summary of the Scale properties*

*** escape352 data ***

Standard deviation of Object parameters	=	2.974
rms estimation error	=	0.668
Separation coefficient (sd/se)	=	4.453
Reliability coefficient (like alpha)	=	0.950

Assuming these represent a population 6 sd's wide, and that bands 3 se's apart are distinguishable, then:

There are up to 9.24 reliably distinct bands of objects.

The final scale spread the portfolios out with a standard deviation of almost 3 units. The average measurement uncertainty for a portfolio was about 0.67 units, and the ratio of these two figures was 4.45. This means that the standard unit of the scale was almost 4.5 times as large as the uncertainty of measurement. This means the portfolios were measured with an uncertainty that is very small compared to the scale as a whole; this ratio is then converted into the traditional reliability statistic – a version of Cronbach's alpha or the KR 20 coefficient.
The value obtained was 0.95, which is very high in GCSE terms. At the moment we do not have much data on reliability in national examination components, though this should change when Ofqual's new programme of sponsored research gets underway. An alternative comparison is with the Verbal and Mathematical Reasoning tests that were developed for 11+ and similar purposes from about 1926 to 1976. These were designed to achieve very high reliability, by minimising inter-marker variability and maximising the homogeneity of the items that made them up. KR20 statistics for those were always between 0.94 and 0.96.
With *escape* in this Design & Technology task a level of internal consistency comparable to those reasoning tests has been achieved without reducing the test to a series of objective items.

Validity

Internal consistency is a component of reliability, and hence of validity. It helps to consider these concepts backwards, in terms of evaluating threats that may reduce reliability and validity, rather than factors that contribute to them.

The principal concern in this *escape* D&T context is that judges may differ in their evaluation of the portfolios they see. Traditionally, marker unreliability like this is divided into three parts: (a) variation in absolute standard; (b) differences in discrimination or spread of marks; (c) differences in conceptualisation of the trait being measured. A principal advantage of the Thurstone method is that the first two of these threats are nullified: the judgements made are merely relative, or ordinal in scaling terms. The interval or ratio aspects of the scale are created out of the entire set of data and in no way depend on the individual judges making interval or ratio judgements.

Only the third threat remains, that two judges might choose a different winner because they differ in what they think 'good' means. The analysis provides reassurance on this in that the very high level of internal consistency achieved means this cannot have been a significant threat in these data. Furthermore, if it ever turns out to be serious then the misfit statistics will discover the source of the inconsistency and allow appropriate compensation to be made.

Moving from reliability to larger concerns of validity, two arguments can be made. The first is that this method is *intrinsically* valid, in that competent judges, of the same sort as are used in examination marking procedures, have given their direct judgements about which of two portfolios is the better, where 'better' means exactly 'of higher quality in terms of what we want to assess'.

Throughout the test development literature examples can be found of test validation that use rank orders as a criterion for validity. Typically, as with the Edinburgh Reading Tests for example, teachers are asked to provide rank orderings of their pupils, and the intra-class correlations of these rankings with test scores is considered to estimate the validity of the test scores. Values of 0.9 or so are considered very strong evidence of validity for the test. The basis for this argument is that the teachers know their students very well, and that the rankings they construct are therefore intrinsically valid. If a test reproduces more or less the same order then it too must be valid. Notice that there are four threats to validity in this argument. First, the teacher may not know their students as well as we assume; second, the students may not perform predictably on the day of the test; third, the test may not

measure the same trait as the teacher judges; fourth, the markers may not be consistent in scoring the test performances.

To validate the whole *escape* D&T assessment we should look for teachers able to give us rank orders of their students that can be correlated with the rankings from the paired comparison analysis of the portfolios. In the meantime, if we consider just the validity of the judgement process itself, then the first, second and third threats described above do not apply, since we would be assessing the validity of assessing the *performances* rather than of the *students*. This validity is essentially measured by the degree of consistency amongst the markers.

Setting grades

Like the raw score scale arising from marking exam papers, the scale on which the portfolio parameters are estimated has little substantive meaning. A suitable linear transformation is needed to support its interpretation. Only a linear transformation will preserve the equal interval nature of the scale the Thurstone method creates but, in principle, any linear transformation will do if it is useful, and the principle is the same however it is done.

For example, suppose a scale of GCSE grades is needed. Two points need to be fixed, and in this case they would be two significant grade boundaries. They might be fixed by looking for portfolios whose quality ofperformance is deemed to be on the borderline between adjacent grades, but this is not advisable when the new component is different in nature from the existing ones. A better approach is to determine, for a defined group of students, how many are expected to pass each of two grade boundaries, probably by counting how many do so on other components.

If, for example, it is decided that 20% of the sample analysed here should get at least an A and 60% at least a C, then these percentages are converted to numbers:

20% of 352 = 70; 60% of 352 = 211

Lines are drawn in the ordered list of parameters below the 70th and 211th students, and the minimum parameter values that get the higher grade are noted:

A/B	Ranks 70/71	2.16
C/D	Ranks 211/212	-0.61

The distance between these two points corresponds to 2 grades, so the width of a single grade is calculated as (2.16 – -0.61)/2, which equals 1.385. We can now define all the grade boundaries by adding/subtracting this number as appropriate:

Grade	Minimum parameter	N	Cumulative N	%	Cumulative %
A*	3.55	39	39	11.08	11.08
A	2.16	31	70	8.81	19.89
B	0.78	64	134	18.18	38.07
C	-0.61	77	211	21.88	59.94
D	-2.00	59	270	16.76	76.70
E	-3.38	42	312	11.93	88.64
F	-4.77	21	333	5.97	94.60
G	-6.15	12	345	3.41	98.01
U	-----	7	352	1.99	100.00

The results of this procedure are marked in the Appendix, and the resulting grade distribution is shown above. The exact positions can, of course, be adjusted a little.

It is worth noting that the average standard error of measurement for a portfolio was 0.668, which is just less than half the width of one of these "GCSE" grades. It is unlikely that many GCSE components – or any that are judged rather than scored quite objectively – could match this level of measurement accuracy.
This illustration simulates GCSE grades, and facilitates aggregating the results with other GCSE components, perhaps by means of a kind of UMS. But any number of grades with any desirable properties could be chosen instead, particularly if the purpose is formative rather than summative.

Summary

Figure 4 summarises the estimation process, with the set of grade boundaries described above added. This illustrates the kind of measurement system the *escape* D&T task is capable of.

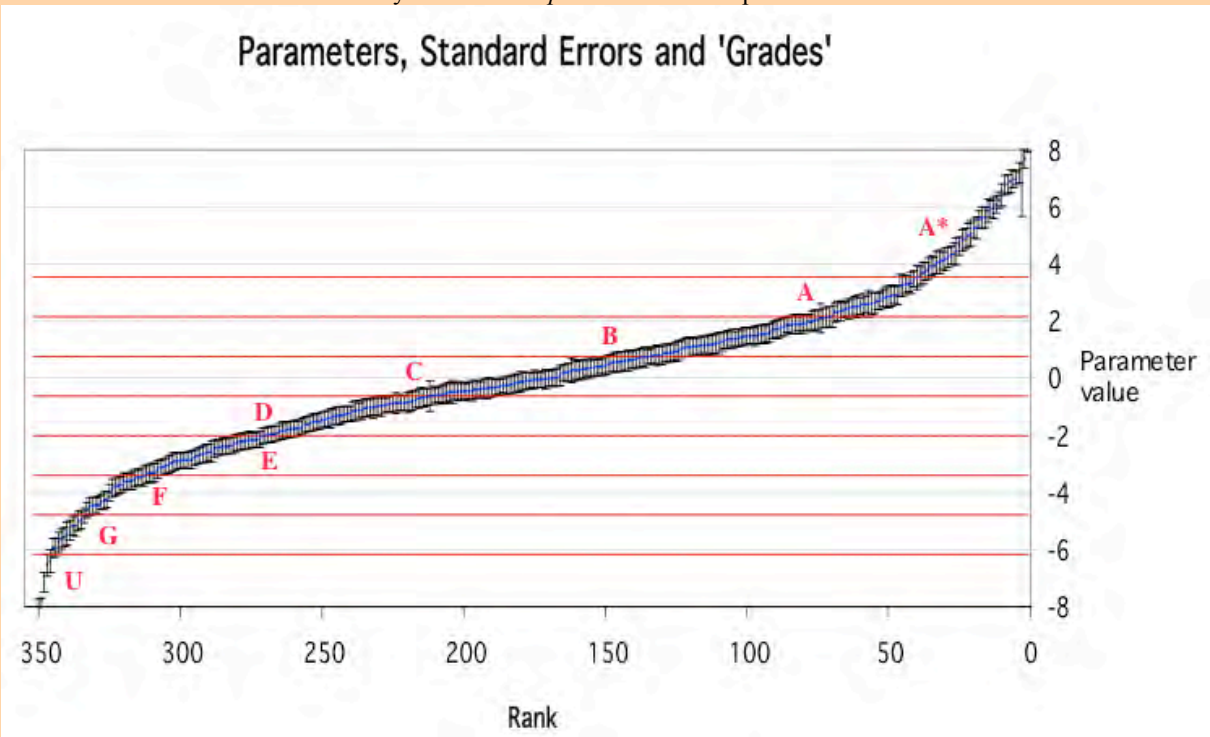


Figure 4: Portfolios and their estimation errors, with added 'grade boundaries'

NB the complete list of 352 portfolios and their associated statistics is sown in Appendix 29

for the benefit of non-statisticians..

For those readers who are less concerned with the details of the statistical analyses in this report from Pollitt, it is worth pulling out – and highlighting - a few of his observations throughout.

Concerning the design of the algorithm in the pairs engine:

In effect, the Swiss tournament system [from the 1st 6 rounds of judging] achieves an approximate targeting at almost no computational cost.

Concerning quality control:

Because every single judgement made can be compared to the outcome predicted (with the benefit of hindsight) from the final rank ordering, very detailed monitoring is possible of the consistency of the judgements made by each judge, and of each portfolio.

Concerning the 'fit' statistic for judges

Theory predicts that this statistic should average 1.00, and in these data it does exactly that.

…the calculation gives 1.64 as a criterion [for fit], and only one judge [out of 28] exceeds this… It may be significant that this judge made only 59 judgements, while the others averaged almost twice as many…

[NB in fact the judge was reluctantly forced to withdraw from the process for personal reasons]

Overall the amount of misfit seems quite acceptable.

Concerning the 'fit' statistic for portfolios

16 [of 352] portfolios exceeded this level, [the acceptable 'fit' statistic] or 4.5%, which is satisfactory for a 5% significance test.

Concerning Pollitt's summary of the process

the portfolios were measured with an uncertainty that is very small compared to the scale as a whole … The value obtained was 0.95, which is very high in GCSE terms.

Values of 0.9 or so are considered very strong evidence of validity for the test. It is worth noting that the average standard error of measurement for a portfolio was 0.668, which is just less than half the width of one of these "GCSE" grades. It is unlikely that many GCSE components – or any that are judged rather than scored quite objectively – could match this level of measurement accuracy.

NB in all the above snippets from Pollitt's report, anything [in square brackets] has been added by us as illuminative detail.

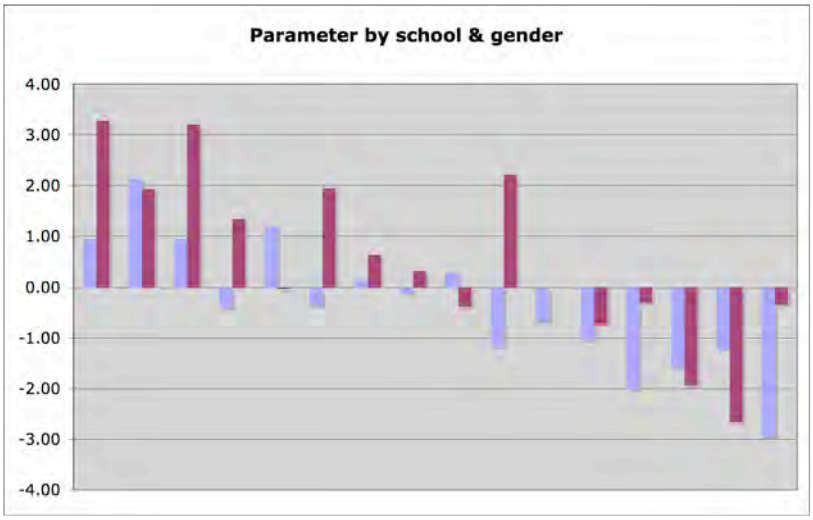
performance distribution through the sample

In section 11 below we describe and analyse the rich diversity of learner performance that is evident in the portfolios. Before getting to that however it is important to examine the performance rank in an attempt to make it meaningful to readers, and we include here three kinds of analysis:

- analysis by school
- analysis by gender
- analysis by 'general ability'

performance by school and gender

When we conducted the APU survey in 1988 (Kimbell et al 1991) we were initially astonished at the 'school effect' in the data. The differences between the mean scores of whole schools were far larger than would have been expected in (say) English or maths. This was (then) a feature of the new-ness of the subject and the differences that can be created in individual schools by teachers with a real understanding of the new field. But interestingly all our subsequent studies have continued to reveal quite significant school effects, and this was certainly true in the e-scape phase 2 testing undertaken in 2006 and reported in 2007 (Kimbell et al 2007).



In the sample analysed this year we include 17 schools; 12 cold schools, 4 hot schools, and one school in which we conducted the very final trials and where the portfolios were sufficiently rich to include in the final analysis. In all these schools the structure of the activity was identical and the resulting portfolios were sufficiently complete to enable them to be effectively integrated into the judging process. Bearing in mind that the full range of parameter scores runs from +8 to -8 the chart here shows the mean

performance of the male and female groups in each school. The overall male mean score was -.4 and the overall female mean score was + .6

A number of interesting things emerge from this chart. The highest performing schools have a mean score of around +2 and the lowest of -2. In some there is a big difference between the gender groups and in others almost no difference. In schools 1 and 3 for example girls outperform boys very substantially, whereas in 2 they are almost identical. Schools 15 and 16 also show inverse positions, with boys doing much better than girls in 15 and girls outperforming boys dramatically in 16. School 10 has the biggest gender difference, but it should be noted that this school only had 5 girls in the sample, three of whom were very high up the rank.

We have speculated on a series of possible reasons for the whole school differences – bearing in mind that these schools are all free volunteers into the system and have all received identical training.

Perhaps … in some schools….

- learners were more familiar with the kinds of concept design challenges that are represented in our task.
- learners were more comfortable / familiar with computer-based kit in a design context
- teachers preparatory work (e.g. the handling collections) provided richer starting points
- the activity was run by the teacher in a more flexible/accommodating manner

performance by 'general ability' and gender

It is clearly important to consider the analysis of performance against some background measure of 'general ability'. This is never a straightforward matter, but for our purposes we asked schools to provide SAT scores in English, maths and science for the test cohort. From these data we calculated three broad ability bands

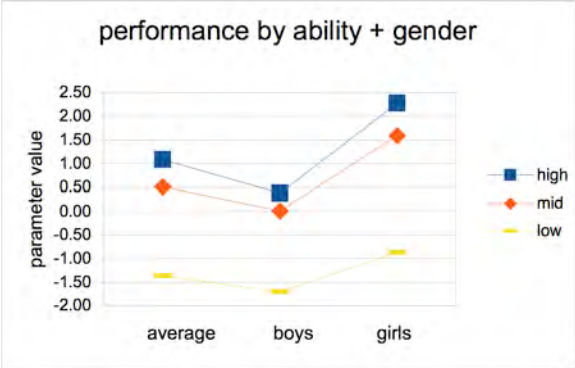
	girls	boys
high ability	37%	63%
mid ability	32%	68%
low ability	40%	60%

The mid-ability band is bigger than the other two, but interestingly, the girls are more-or-less evenly distributed through the sample.

We then calculated the mean parameter value score for the three bands, split the bands by gender and calculated the gender mean scores. The charts below shows this distribution

	ave	boys	girls
high	1.08	0.38	2.28
mid	0.51	0.01	1.59
low	-1.36	-1.7	-0.87

Presenting these data in chart form exposes two interesting and consistent trends. First the ability bands show almost exactly parallel gender differences. The girls high/mid/low performance is exactly in line with (and better than) boys high/mid/low performance. Second, the mid-ability band is NOT positioned midway between the high and low performance bands. In fact, the ***difference*** between the performance of the mid-ability and low-ability bands is approximately ***four times bigger*** than the performance difference between the mid and high ability bands.



Another way of saying this would be that somehow the mid ability band is performing almost on par with the high ability band – and well above the low ability band. And this trend is independent of gender.

In order to inform this matter – as well as the general disparity of performance across schools, we examined some of the observation data. These data were collected by the research team as the activities were underway, but as a consequence of the trouble-shooting that was occasionally required of us, the data sets are not always complete. Accordingly it is difficult to identify clear trends in these data.

However there is one issue that bears significantly on the performance data reported here. It concerns the 'engagement' of the learners with their task. We have reported this on a sliding scale from:

Stationary = no movement = off task = disengaged
Poddling = in 'tick-over' mode = doing enough but not exerting
Motoring = working flat out = fully engaged and progressing

This classification originated in an ESRC research project conducted by the TERU team in 1994 and has been used several times since then. For a full account of this work see (Kimbell Stables and Green 1996)

If we examine the 'motoring' data from the e-scape observation forms an interesting trend is clear – and perhaps not surprising. The schools where there is a high level of 'motoring' (in 80-90% of recorded incidences) are those at the upper end of the schools' performance distribution. The school with the lowest level of motoring (46% of recorded incidences) is at the bottom of the schools' performance distribution.

It would seem that at least one of the significant performance variables between schools is the extent to which teachers have been able to create the conditions in which their learner group takes the activity by storm … 'motoring' through it with a high level of engagement. And this is not just a matter of being able to switch on a group of committed, high ability learners, for the ability data suggests that – proportionately – the bigger mid ability band has done very much better than might be expected. Given that the activity lasts for 6 hrs (two 3hr mornings) the extent to which the best performing schools have maintained this high level of engagement throughout the activity is remarkable, and reflects well on the management by the teachers as much as on the performance of the learners. By lunchtime on the 2nd morning, the learner groups (and the teacher) were in need of a break.

judging the science portfolios

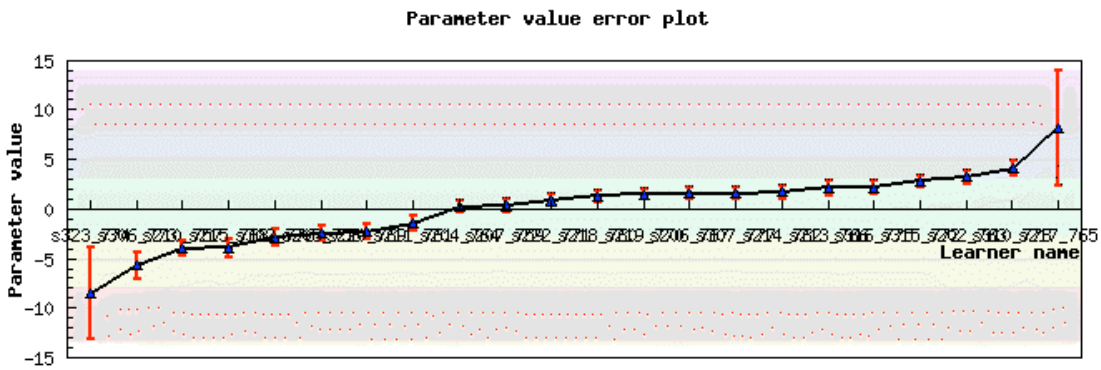
For the science portfolios, judges were recruited from the Bath schools in which the trials had taken place, and from recommendation by the PGCE secondary science tutor at Bath Spa University. The judging panel was:

- Dr Frank Martinelli – Ralph Allen School, Bath
- Jo Tidball – St Gregory's Catholic College, Bath
- Nicky Foster – St Gregory's Catholic College, Bath
- Naresh Claire – John Bentley School, Calne, Wiltshire
- Dave Harries – John Bentley School, Calne Wiltshire
- Dan Davies – Bath Spa University

The team were to judge 22 portfolios, drawn from the last two pilot sessions (since the Birley Community College trial was only 2 hours in duration, the portfolios arising from this were incomplete and could not be used in the judging process). The judges all attended a training day on 4th November 2008 at Bath Spa University, with a programme that was loosely based on the d&t judge training session. The science training programme was as follows and judges subsequently reported that the training had prepared them well to undertake the judging process.

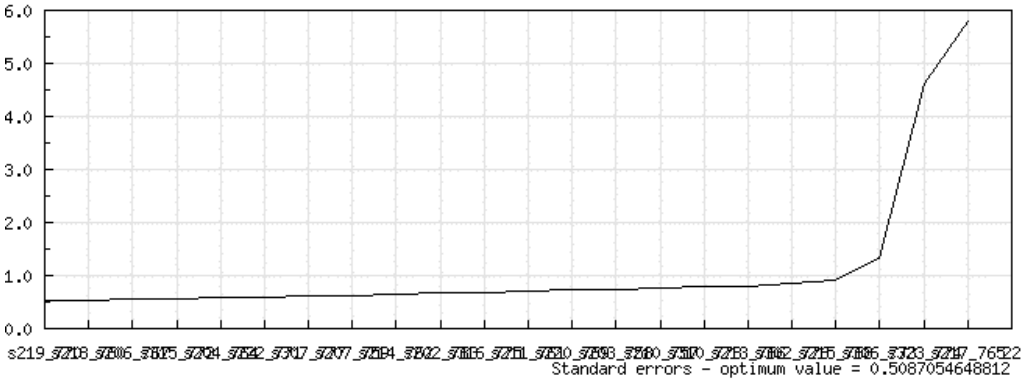
10.30 Introduction: E-scape:
The process: theory and practice (Richard Kimbell)
11.15 The Science Road Safety activity (Dan Davies, Frank Martinelli)
11.45 Identifying and assessing 'good science' (Discussion)
12.30 Lunch break
1.00 Judging:
the 'Comparative Pairs' methodology (Richard Kimbell)
2.30 Accessing the pupils' work;
the e-scape on-line management system (Dan Davies)
3.45 Next steps: schedule, evaluation of methodology, data analysis,
Awarding Bodies

The judges started the online judging process during the training session, completing all 220 judgements (20 rounds, 37 judgements each) over the next four weeks. We have reported at length judges' reactions to the process in section 9 above, based principally on the questionnaire data gathered from them at the end of ther process. They reported that although the first few judgements were difficult to make and required a lot of note making and moving backwards and forwards between portfolios, once the system started 'chaining' the process became much more straightforward, and they liked being able to compare work from different schools. Another advantage of this holistic approach to assessment was seen as preventing teachers from teaching to the test as much, thus giving candidates a fairer chance. The idea of pairs comparison felt strange to the judges at first, but most felt in retrospect that it had worked well and had been surprisingly easy. They were enthusiastic about using judging as part of general science assessment practice and felt that students could potentially improve their own performance by being involved in the judging process. At the end of this process (after round 20) the parameter value error plot was as follows:



The above chart indicates that the errors on judging 19 of the 22 portfolios were acceptably small, suggesting a high level of reliability. Only the top and bottom portfolios (ranks 1 and 22) produced large errors, which is a feature of the system rather than the reliability of marking (rank 1 consistently 'won' all its pairs comparisons with all judges, whilst portfolio 22 consistently 'lost' them all). Looking at the errors for each portfolio (see below) confirms this picture. The figure for 'Standard errors – optimum value' is what would be expected after round 20 of the process. As can be seen, 19 of the portfolios have an error of less than 1.0, with the majority close to the 'optimum' of 0.5. Only the two 'outliers' (ranks 1 and 22), together with rank 21 to a lesser extent, lie outside this acceptable figure.

Standard errors on judgement after 20 rounds:



The 'misfit' figures for 5 of the 6 judges are also low, ranging from 0.03 to 0.55, indicating a strong level of agreement between them. Interestingly, the only judge with a large 'misfit' figure (1.39) was Prof Davies - the developer and administrator of the activity! This may indicate a difference between the perceptions of a researcher with a primary science background and a group of practising secondary teachers.

Simulated Grade Awarding

This took place between the Prof Davies and Frank Martinelli, a senior examiner for AQA science, on 9th December 2008. They first reviewed the judging process, agreeing that performance generally by pupils was not as high as would be expected in a GCSE examination, for the following reasons:

1. The majority of portfolios in the judging process were from Y9 pupils – two years before they would normally undertake GCSE.
2. Pupils had been given no preparation for the test, and we had been at pains to stress that we were 'not testing them', therefore they may not have taken it as seriously as a GCSE examination.
3. The novelty of the Ameos and unfamiliarity with the scientific equipment gave an impression of 'fun' and encouraged experimentation, which may have detracted from the quality of some answers.
4. Some students ran out of time on some boxes, and some data was not captured to the online portfolios.

However, when reviewing the judging history for each portfolio in the rank order, the degree of agreement between judges was noted, and the identification of key discriminating criteria, were listed as follows:

- General level of scientific explanation
- Interpretation of data/ choice of graph
- Understanding of technical terms (e.g. accuracy, precision)
- Choice of independent variable for their own investigation
- Perseverance in collecting data
- Reasoning with data
- Engagement with equation
- Ideas for testing solutions

In order to generate descriptors for different levels of performance 'grades' it was decided to summarise the judges' comments on each portfolio in terms of performance against the above criteria. Each was rated on a scale of zero (no engagement, data missing) to 3 (good), then aggregated the scores for each portfolio in the rank order using this atomised approach (see table below).

Outcome of Grade Awarding Meeting 9.12.08

Portfolio Rank Order	Gender	Scientific explanation	Interpretation of data/ choice of graph	Technical terms (accuracy, precision)	Choice of independent variable	Perseverance in collecting data	Reasoning with data	Engagement with equation	Ideas for testing solutions	**Total**	**Tentative grade**	**Y9 Science SAT**	**Predicted Grade**
1	M	3	3	1	1	3	3	2	3	**19**	**B**	**7**	**A***
2	M	2	3	1	1	3	2	0	2	**14**	**B**	**7**	**A**
3	M	3	3	1	3	3	2	1	1	**17**	**B**	**7**	**A**
4	F	2	3	1	1	3	2	2	2	**16**	**B**	**6**	**B**
5	M	2	2	2	3	1	1	0	2	**13**	**B**	**7**	**A***
6	F	2	2	3	3	2	1	0	2	**15**	**B**	**6**	**B**
7	F	2	2	2	1	2	1	0	2	**12**	**C**	**7**	**A***
8	M	2	2	2	1	1	1	0	2	**11**	**C**	**7**	**A**
9	M	2	2	2	3	2	2	0	0	**13**	**C**	**7**	**A**
10	M	2	1	2	1	2	2	0	1	**11**	**C**	**7**	**A***
11	F	2	2	2	2	1	1	0	2	**12**	**C**	**6**	**B**
12	F	2	2	1	2	1	1	2	1	**12**	**C**	**7**	**A**
13	M	2	1	2	1	2	1	1	1	**11**	**C**	**6**	**B**
14	-	2	1	1	1	2	1	1	1	**10**	**C**	**-**	**-**
15	M	1	1	1	1	2	1	0	1	**8**	**D**	**7**	**A**
16	-	1	1	1	1	1	1	1	1	**8**	**D**	**-**	**-**
17	M	1	1	1	1	0	1	1	2	**8**	**D**	**5**	**C**
18	M	1	2	0	1	0	0	0	1	**5**	**E**	**5**	**B**
19	F	1	1	0	0	0	0	1	1	**4**	**E**	**7**	**A**
20	F	1	0	1	0	0	0	1	1	**4**	**E**	**6**	**A**
21	M	1	0	0	0	1	0	0	1	**3**	**F**	**6**	**C**
22	M	0	0	1	0	0	0	0	1	**2**	**F**	**6**	**B**

The Spearman (ordinal) correlation between these aggregated scores and the rank order was high at 0.964, though it should be borne in mind that they worked down the rank order in this exercise.

The aggregated scores appeared to fall into five levels of performance, which were given the 'tentative' grades B to F, since no portfolio provided evidence of 'A' or 'A*' performance in GCSE terms, and the lowest two in the rank order were very weak. These tentative grades were then compared against the Y9 SAT results and predicted GCSE grades for the students concerned (data for ranks 14 and 16 are missing as the students concerned did not write their names on the Ameos).

The Pearson (interval by interval) correlations between the data emerging from the judging and grade awarding process and data from the schools were as follows:

Between	And	Pearson's R	Significance
Rank order	SATS score	0.507	0.02
Rank order	Predicted GCSE grade	0.537	0.02
E-scape tentative grade	SATS score	0.431	0.06
E-scape tentative grade	Predicted GCSE grade	0.467	0.04

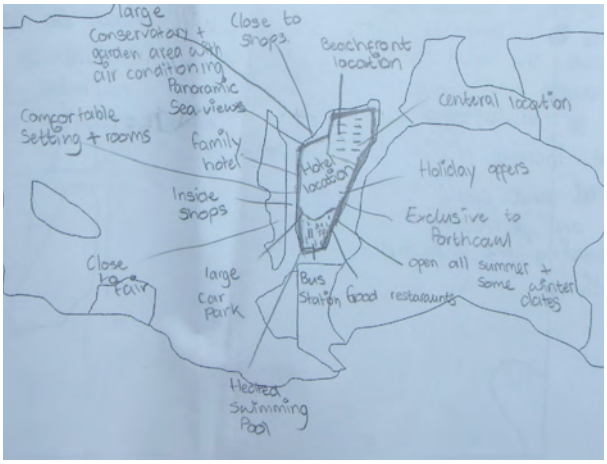
As can be seen from the table above, the correlations between the rank order generated by the pairs comparison process and the school SATS and predicted GCSE grades were slightly stronger (and more significant) than those between the tentative grades we had assigned them during the awarding process.

The full research report from the science team is included as Appendix 19

judging the geography portfolios

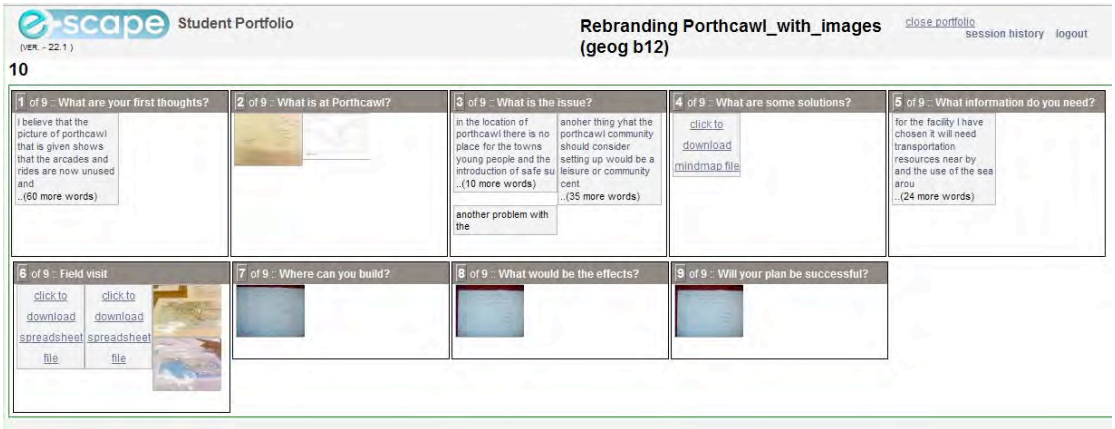
In September 2008, a team of 8 judges met and were trained to assess the work done by the 61 students in the three trial schools. The judges were drawn from a range of experienced geography teachers, some with direct and current GCSE Examiner experiences. The assessments were completed by 7th November, the date on which a further meeting was held to review what had been done. Pollitt, also provided a summary of the raw data and accompanying analysis (see Appendix 28)

It is important when looking at the quality of work produced by the students in the trials that something of the background is understood. All the trials were conducted as 'one-off' sessions that were completely out of the context of the students' normal sequence of work. This was not ideal and would not represent normal practice as it meant that they often did not have the knowledge and understanding that would have allowed them to achieve the highest standard of work of which they might otherwise have been capable. The trials were in effect, assessments of their underlying abilities to think geographically, but without the foundation of specific knowledge and understanding that would have more fully informed their ideas. The timing of the trials is also likely to have affected the students' performance. The students were all towards the end of year 10, their first year of their GCSE course.



Annotation on a hard copy map to give reasons for the choice of a site for a new leisure development in Porthcawl. The map is then photographed using the PDA.

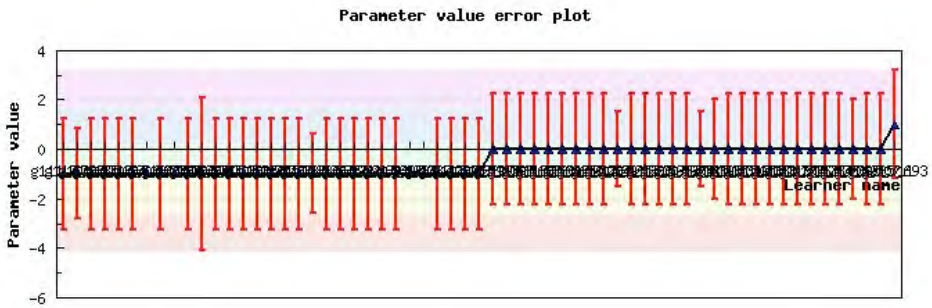
The assessment was done online using the e-scape system that became available in September 2008. This provided the judges with a managed system in which the work of pairs of students were automatically selected for comparative marking. All files in each student's assessment portfolio were made available. Each judge worked through the portfolios, making judgements between portfolio A and B, each time giving a 'winner' of the pair. There was an opportunity to make summary notes, though the scripts could not be annotated as is possible with hard copy work. The result of the process was a set of statistics and graphs that provided a rank order for all 61 students in the three trial schools.

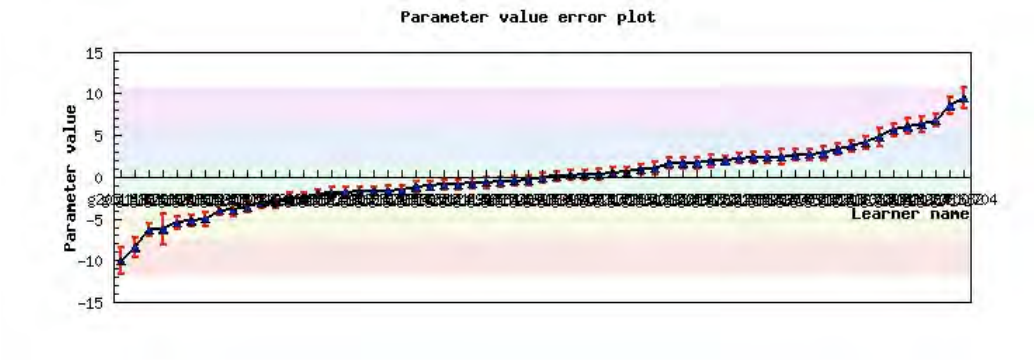


An online portfolio for one student, consisting of text images and sound files

Statistics for the cohort of 61 students showed that the resulting rank order would be acceptable in terms of its accuracy and reliability. This is indicated by the 'misfit figure' that needs to fall within +/- 2. Of the 8 participating judges, 7 were within this range, 2 being within +/1 .5.

It is noticeable how the degree of accuracy increases during each progressive round of marking, culminating in the final result. This is shown by the parameter value error plots. The first round plots showed a large error value (indicated visually by the length of the vertical bars). By the final round (round 17), this had been considerably reduced, indicating that there was a high degree of agreement between the judges.





Pollitt concluded his analysis of the geography judging as follows:

> "We can conclude here, then, that the judging was highly consistent across judges. A more indirect interpretation is that we can calculate a 'reliability coefficient' that is comparable to the classical alpha, or KR20, coefficient often reported for published tests. The value here is 0.96 which is extremely high, meaning that these students' pieces of work have been graded successfully, and we can have considerable confidence in the rank order obtained. It is unlikely that any real GCSE component achieves marking as reliable as this."

The judges did not attempt to allocate a GCSE grade to any of the work as this was not part of their brief. It would appear from the statistics, however, that this would be possible with some further work. As Pollitt indicates..

> 'One consequence that follows directly is that these data would allow us to define at least 10 distinct grade levels of quality amongst these pieces of work, more than enough for this to operate as a GCSE component, for example.'

It can, however, be noted that the emphasis in the trial work was on assessing the principles of the process and in evaluating the assessment tools, rather than regarding marking for accuracy as the focus.

Judgement history for Fred Martin

Start time	End time	Duration (hours:mins:secs)	Portfolio A	Portfolio B	Judgement notes	Winner
23 Oct '08 11:05	23 Oct '08 11:09	00:04:18	Fairly basic ideas throughout, not showing enough depth of thought or analysis of the resources - perhaps a grade D/C. (662) -	Some sound ideas, but nothing stands out. Some Tasks misunderstood and some muddled answers. (670) -	A marginal decision, both perhaps C grade.	Portfolio B
23 Oct '08 11:02	23 Oct '08 11:04	00:02:21	Some good ideas, but some Tasks misunderstood. . (689) -	A genrally sound set of answers, though of D/C depth and quality. (710) -	Fairly close - a D with C potential.	Portfolio B
23 Oct '08 11:00	23 Oct '08 11:02	00:01:51	Generally poor quality work with some being unreadable. (671) -	B is slightly more focused in responses compared to A, but there is little between them. B lacks a map to show the location. (692) -	B seems better overall - it is more complete.	Portfolio B

Online data about markers, giving details of time taken and notes about each student's answers

The full research report from the geography team is included as Appendix 18

11. characterising performance in the d&t portfolios

an approach to characterising capability

In phase 3 the portfolios include a wider range of media (e.g. video, enhanced photography) than had been the case in phase 2. The resulting 350 multi-media portfolios represent a fantastically rich resource for exploring learner capability through web-portfolios. What follows in this section is an attempt to capture the essence of these portfolios. It has not been possible to analyse the contents of all of them. That would be a monumental task. Rather, we have used a sampling procedure to take slices through the layers of capability revealed by the ranking process. Moreover within the sampled slices, we have examined the work in two ways, one focusing on the *product*, or outcome, and the other in a more holistic way on the *process of designing* used by the student.

In the first instance the top 40 portfolios were split into two alternately ranked groups (odds and evens), with one group being analysed in terms of product by one researcher, and the other in terms of design process by another researcher. From this, a range of performance indicators were drafted and agreed by both researchers. The next step was to look at the bottom-ranked portfolios in a similar manner. Finally portfolios were chosen from the middle-range (upper middle and lower middle) and especially those that were identified by judges in their 'tell Richard' comments. In all around 33 portfolios were examined in detail in terms of product, and 33 in detail in terms of learners' designing approach.

The final step was to look across the analysis grids to try to identify patterns or trends in order to arrive at a point where common statements could be made to describe progression and performance criteria across the rank. These statements were then compared to relevant GCSE d&t grade criteria.

We begin this section by examining just one of the boxes in the portfolio – the 'inspiration pitch' in which learners record where their ideas came from. It provides an interesting opening since one of our key ambitions with this work has been to encourage the fluency and richness of learners' ideas.

the inspiration pitch: where do ideas come from?

A key aim of the e-scape activities is to enable all learners to demonstrate their ability to be creative and innovative. As a way of 'drilling down' into the data to explore this aim, the inspiration pitch has been analysed at some length as it has a unique place in the activity in offering the learners an opportunity to explain the source of their inspiration.

Introducing the inspiration pitch – when, why, how

The 'inspiration pitch' was a new addition to the activity structure in Phase 3. Its purpose was to gain insight into where the learners ideas had come from, particularly in the light of elements we had introduced into the activity specifically aimed at inspiring them to be creative. The 'pitch' was in the form of a voice file – drawing on the success

of the final pitch at the end of the whole activity that was introduced in Phase 2. It was placed at the end of day one as, by this stage, we anticipated that learners would have a reasonably developed concept for their pill dispenser. The pitch was introduced in the following way.

"There is one final task for this morning. As we have said from the start, we are very interested in your design ideas – and we would like to know what inspired them. What prompted you to develop your idea as it is …. and what is it you are trying to do in your design. When we move on to screen 15, click the link to the sound file and record a 30 sec sound-bite explaining where your ideas came from, and how you hope they will develop".
(Administrator's Script)

How we had aimed to 'build in' inspiration

Through the Assessing Design Innovation project (see section 1) we had developed the concept of an 'inspiration' collection. This was a carefully selected collection of objects (a handling collection) that were chosen for the potential they had to:

- *Inspire learners to explore the unexpected, novel and provocative; to think the unthinkable and step outside the box.*
- *Help learners unstitch a concept from an existing product that they can apply in an entirely new and innovative way to their own design*
- *Provide a source for creative inspiration throughout the task*

(e-scape training materials)

Contents of a typical handling collection

We had found previously that learners responded well to such a collection and that, in the better work, the products in collections had a clear, innovatory impact on learners' developing ideas. In addition, and based on research as far back as the APU project (Kimbell et al., 1991), we were mindful of the need to create a challenging design brief – a task that was 'issues rich' and that would provide the impetus for learners to understand the needs of the clients they were designing for and for these needs to be the 'jumping off' points for creative responses.

Client card for the Pill organiser activity

The APU project had developed in us an acute awareness of the importance of context and so, to enable the learners to think themselves into the task in as immediate way as possible, we created a range of 4 different, credible 'clients' – a young boy with an allergy condition, a teenage migraine sufferer, an adult who engaged in active sports and an elderly woman who was both forgetful and confused about which pills she had

to take when. For each client, we developed a 'user profile' – and created a set of four profile cards for learners to refer to in identifying and specifying their own client.

The range of things identified

The general overview of their responses when asked about the inspiration for their ideas is shown in Chart 1 – which indicates the number of incidences when a particular category of inspiration was mentioned. The 1st two columns on Chart 1 it can be seen that, in total, 14% of learners didn't make an inspiration pitch at all – and this is dealt with further on in this section. For the rest, the chart shows the number of times learners mentioned a particular category of influence – from "don't know what inspired me" to "it came from my head". As we had anticipated, the handling collections were identified the most. But what surprised us were the number who referred to the user profile cards and also those who cited some external source.

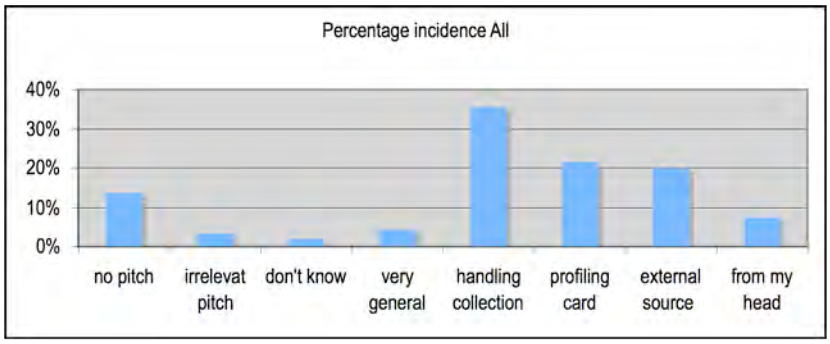


Chart 1: An overview of what learners were inspired by.

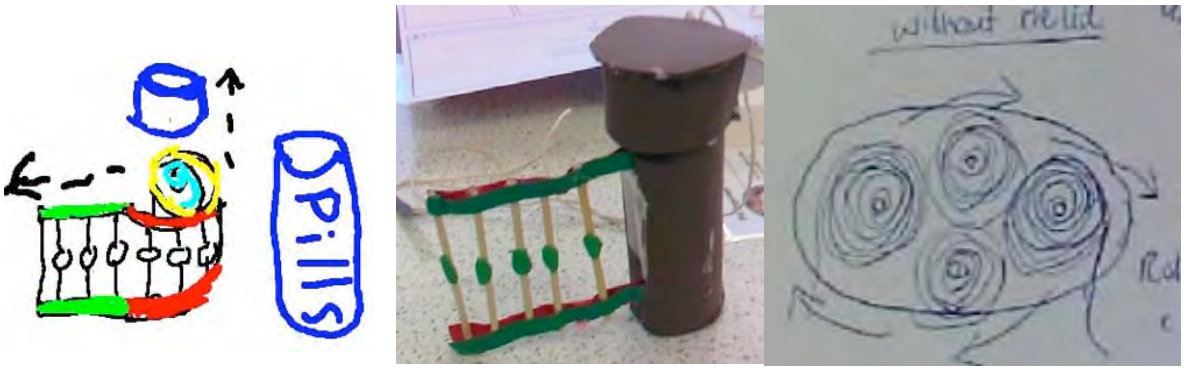
A number of learners mentioned items that fell into more than one category. In what immediately follows the chart, we illustrate the range of ways learners identified their inspiration and illustrate how this impacted on their developing ideas.

Specific mention of the handling collection

For some learners, their whole design was influenced by one product in the handling collection, as shown in Figure 1. In this instance, the teacher had provided a very diverse range of objects and the learner saw his opportunity in taking an idea from a sheet of resistors. He went on to develop a pill dispenser that contained sheets of pills that fed out of a container by turning a handle.

"Well the thing that inspired me was the link of resistors joined together. I thought that would make a good roll of pills."
(Pupil No 02/12: ranked 33)

Figure 1 The inspiration and developing idea of Pupil 02/12



Other learners saw how they could adopt concepts from different objects and combine them to develop their own idea. In the example shown in Figure 2, the learner has extrapolated concepts from 4 different products to exploit in his own design.

Figure 2 The developing idea of Pupil 11/14, drawing on multiple

"What inspired me was the objects lying around. Certain designs were inspired specifically by this thing on the table, which smelt quite nice and had a lid that spun round. That inspired my first design - I completely copied that in a way. The other one was a belt buckle and a watch, which inspired - there was a belt lying around, you know. And a watch and stuff like that. Another thing was this thing that you flick up in the air and catch a ball."
(Pupil No 11/14: ranked 31)

Unstitching a concept

In the training material, we had emphasised the potential of a good handling collection to include products from which learners could 'un-stitch' a concept – and this was illustrated nicely by one of the Primary aged children in the trial, in describing how he took his idea from a particular design for an eraser dispenser.

> *"Where I got my idea from is a rubber and it span round. Well you pull this little out-ridge and it comes round and then the rubber gets shown, and I thought, wait, couldn't you do that with pills. Instead of it already being connected Its scooping them up and showing you them. I thought that was a pretty good inspiration."* (Pupil No. 13/02: ranked 102)

Making good use of the profiling cards

The profiling cards were a major source of inspiration, as learners used them to consider the needs of their clients, as is illustrated by the following two examples; one designing for the elderly and one for a child. What is impressive with both is the way in which the initial inspiration is tenaciously held on to as the very first visualising of the idea (drawn in the first 5 minutes directly onto the Ameo screen) is developed through to a working prototype.

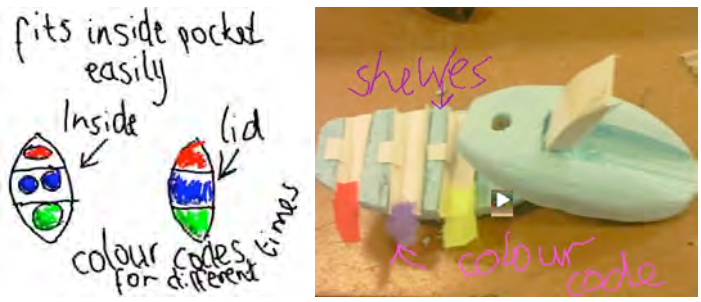


Figure 3 designing a pill dispenser for the elderly

"I based my design on one of the cards that told us which people would be using the pill dispenser. I read about an elderly lady who got confused about which pills to take, and when. I tried to make my design simple, aesthetic and compact."
(Pupil No 15/08: ranked 53)

"I was inspired by one of the typical users needs who says he loves cartoons and going to the zoo. As I read the zoo, animals started going through my mind and I was seeing what children would find as entertaining. I thought of a cow pooing out a pill."
(Pupil No 15/20: ranked 59)

For some learners, the profile cards caused them to think of particular clients as they personalised the profile to someone they new – for example a younger brother who didn't like to take medicine, or an older relative who had similarities to 'Alice' the elderly woman on one of the cards. These learners often cited the inspiration as coming from this personal experience

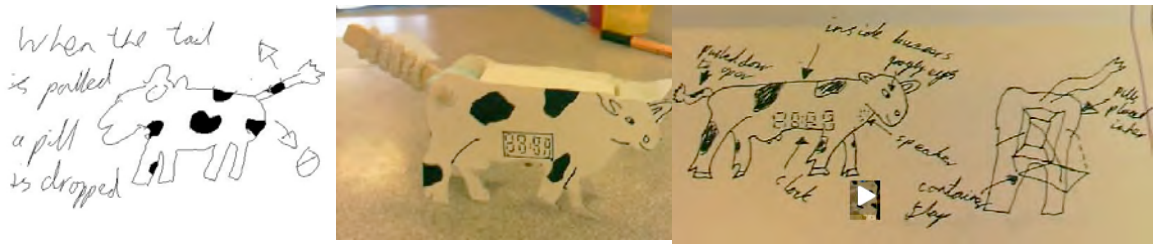


Figure 4 The developing idea of Pupil 15/20, designing a pill dispenser for a child

.

> *"The thing that inspired me was that my granddad takes lots of pills so if I could create one this maybe would help him take it and not forget in the evening or the morning, forget to take them which would be very vital to his health. He has been a big role model in me creating this product."* (Pupil No 11/10: ranked 70)

Drawing from an external source

20% of learners took their inspiration either wholly or in part from external sources – things from outside the classroom. This ranged from products they had at home, products they had experienced earlier in their lives, and products they had seen used in unusual ways – as is the case with Pupil 08/06 who took his inspiration from a James Bond movie.

> *"My design ideas came from a movie and a sequence of films. One was James Bond. I'd seen this necklace - it was containing diamonds and so I thought the diamonds could be pills and that could work. Another was Prison Break and you put little messages in."* (Pupil No 08/06: ranked 21)

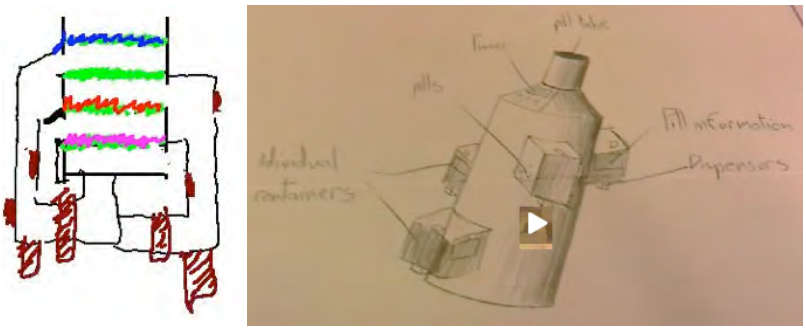
This learner was designing for a teenager who didn't want to draw attention to the fact that they always had to carry medication with them and went on to design a pill dispenser disguised as a fountain pen. As with the use of the collections brought into the classroom, for some, this involved drawing together inspiration from more than one product – as was the case with a second Primary learner who took:

"ideas from a dog lead and I saw loads of people at Glastonbury wearing things that were pretty cool and I thought that I could turn it into a pill dispenser. So yeh, I got it from a dog lead and Glastonbury. Thank you very much. Goodbye."
(Pupil No. 13/15: ranked 72)

Figure 5 The developing idea of Pupil 13/15, combining ideas from an external source

'Just popped into my head'

A significant minority of learners (8%) reported that ideas just 'popped' into their heads. For some, there was evidence that they were talking about what could be termed a 'creative leap' and the initial inspiration then led to purposeful development towards an imaginative solution. For others, the statement had more negative overtones – implying almost a lack of inspiration. The contrast between these two positions is illustrated in the following two examples: the first from a learner ranked 42nd overall, and receiving an A Grade; the second from a learner ranked 304th, who received an E Grade.



"To tell you the truth I have no idea where this idea came from. It just came into my head. Things do that -they just come into my head from nowhere. I don't know why. But it was a good idea, so I kept it."
(Pupil No 10/14; ranked 42)

Figure 6 The developing idea that 'popped into the head of Pupil 10/14

"It just came off my head, I just got an idea, yeh. Nothing much."
(Pupil No 12/10: ranked 304)

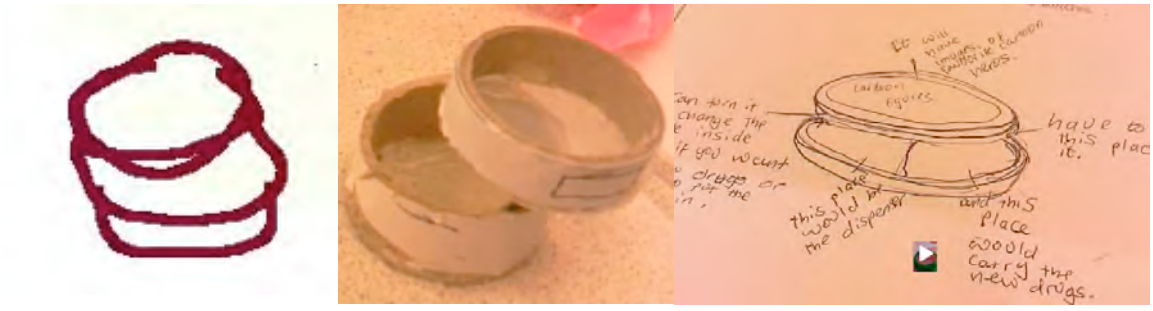


Figure 7 The developing idea that 'popped into the head of Pupil 12/10

The relationship between inspiration and gender, performance and school

Looking more deeply into the data, it is apparent that the ways in which learners report what inspired them is not evenly spread. There are clear trends showing relationships

between what was drawn on and the learners' performance. Exploring this further shows a link to what we described earlier as the 'school effect'.

Interestingly, in exploring these relationships, the one area where there was very little difference was gender. This lack of difference is an interesting finding in itself. There are, however, two small differences; first that boys appear slightly more likely to say that inspiration just came into their heads and second that girls drew more inspiration from the user profiling cards. This difference was quite marked (9%) and resonates with research more generally that indicates that girls are more likely to be 'user' focused. (Kimbell, et al., 1991; Kimbell, Stables & Green, 1996)

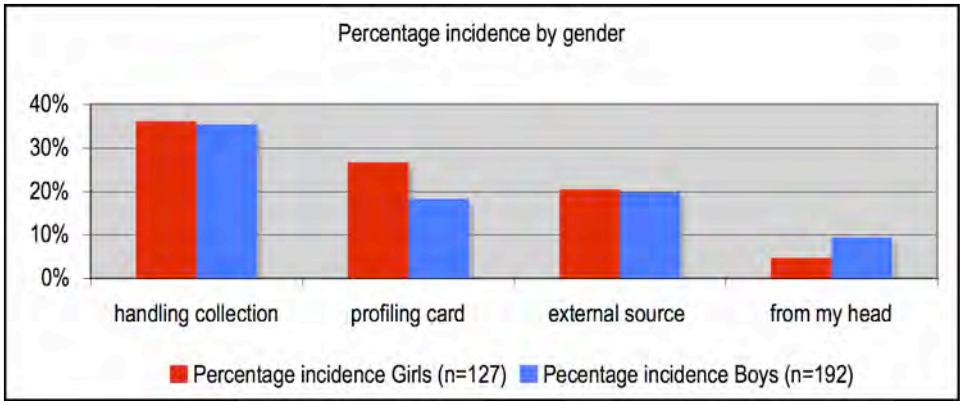


Chart 2: Gender differences.

The performance effect

However, there are more noticeable differences when comparing the range and amount of elements referred to in relation to performance. Chart 3 shows the incidence of categories being cited by learners grouped by the grades received. (Some grades have been clustered to facilitate more even sample sizes). Certain things are immediately apparent. First, the higher the performance, the more likely that the handling collections will have some impact. It is interesting that the focus on the handling collections is particularly strong at Grade C – and this might be explained by the concrete nature of handling collections, and the way they might afford a bright, but less academic learner a very immediate way of identifying a useful design concept. The chart also indicates the very strong use that the highest performers have made of the profile cards - more than twice any other grade. There are also indications that they were using these in conjunction with the handling collections. Finally, it would appear that the lower performers are more likely to have drawn inspiration from inside their own heads – there being very little evidence of this in Grades A*, A and B, and where this did occur, it was more likely to be of the 'creative leap' nature indicated in Figure 6 above.

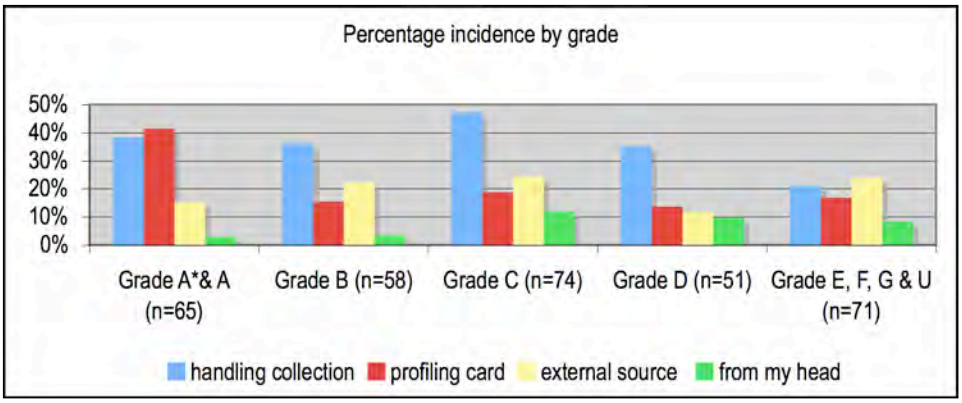


Chart 3: The impact on performance.

The school effect

When analysing the data to compare differences between schools, a number of aspects become apparent. Chart 4 shows the range and amount of references for each school, and is organised by average performance of each school – School 08 on the left having the highest mean performance, School 12 on the right having the lowest mean performance.

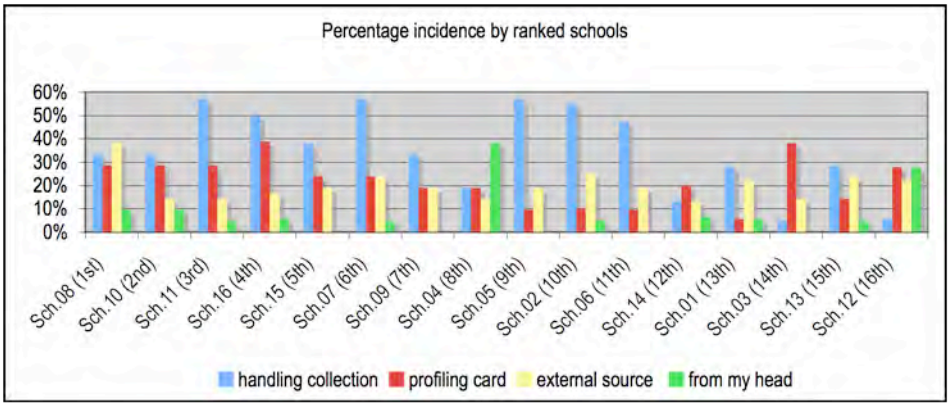


Chart 4: The 'school effect'

While there is noticeable variation, there are still clear indications that drawing on the handling collections is likely to have a positive impact on performance, and a trend to suggest that where this is balanced with drawing on the profile cards and from external sources, performance is likely to be positively affected. This is demonstrated most forcefully in School 8, which had the highest mean performance in the trial and a very balanced reference to handling collection / user profiles / external sources.

But there are some noticeable examples that 'buck' the trends – and in particular School 04 (a mid to high performing school) that made very limited reference to the handling collection; School 02 (a mid to low performing school) that made considerable reference to the handling collections; and School 03 (a low performing school) that made minimal reference to the handling collections and considerable reference to the profile cards.

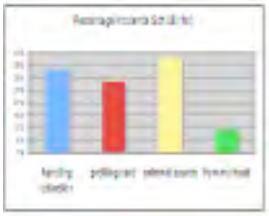
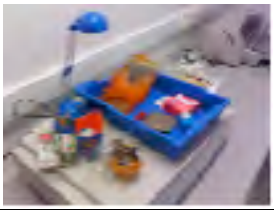
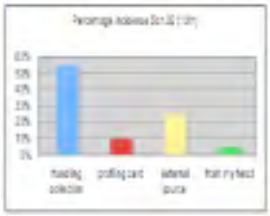
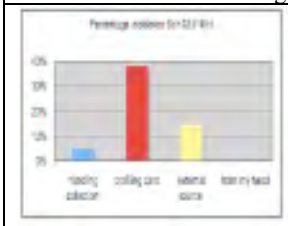
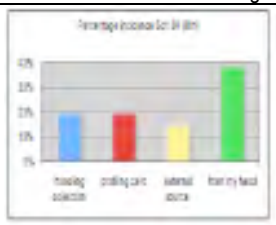
Sch.08 high performance (ranked 1st) Balanced use of handling collection	School 02 mid-low performance (ranked 10th), Very strong use of handling collection
 Comments from observation sheet "Delightful explanation at inspiration objects raising lots of design issues (wish I'd had a video). Pupils desperate to explore collections. Terry's chocolate orange / cd rack"  	 Comments from observation sheet "quite teacherly - went beyond script but stressed ideas. Introduction of task is clear" 
Sch.03 low performance (ranked 14th) Limited use of handling collection	**School 04 mid-high performance (8th) Limited use of handling collection**
 Comment on Observation sheet "Inspiration collection ok, but no individual handling collection." 	 Comments from observation sheet "this bit seemed to come as a bit of a surprise to [teacher] ... [learners] skipped [exploring] inspiration collection to pick up time" Add image Add image

Table 1 Comparing the use of handling collections across 4 schools

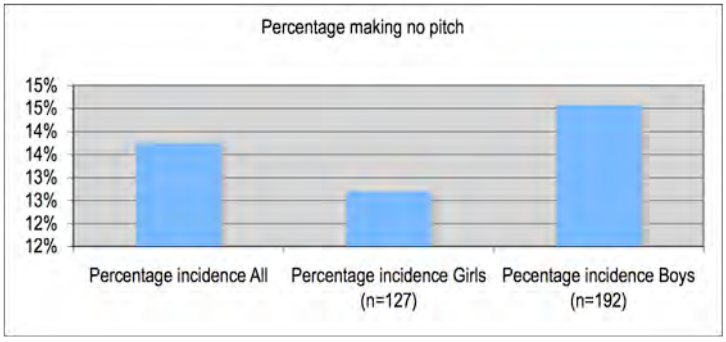
Looking in more detail at the data collected by the researchers observing the trials, it appears that there are two particular aspects that have impacted on the value of the handling collection – the quality of the collection itself, and the way it was introduced to the learners. The variations witnessed were despite the fact that training support was provided to the teachers running the activities - guidance on both what to include in the collection and how to present it during the activity. Table 1 shows the contrast between high and low performance and also between the two mid performing schools.

It is apparent in these examples that in the two schools where handling collections were drawn on quite extensively (School 02 and School 08), the collections were introduced to the learners in a fulsome and enthusiastic manner, whereas in the two where less use was made, the introduction was more minimal. In addition, from the photographic evidence of the actual collections provided it can be seen that the collections in School 02 and School 08 were both more extensive and more varied than in the other two schools.

To pitch or not to pitch

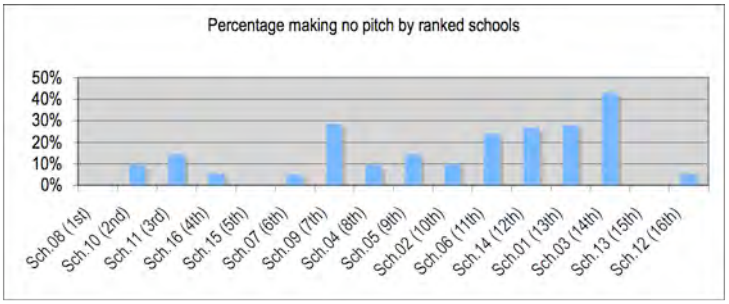
In 14% of the portfolios there was no evidence of an inspiration pitch. In most cases we don't know whether this was due to a failing in the technology or to the learners

actively choosing not to make a pitch (although in certain instances it is clearly the latter as the learners chose to record a 'pitchless' file.)



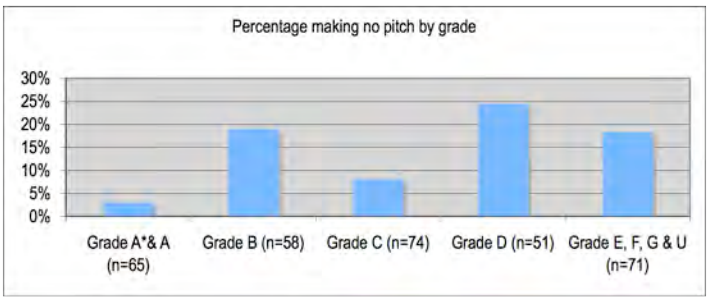
But analysing the incidence of 'no pitch' does still provide some insights. First, the boys appear to be slightly more likely to have made no pitch than the girls.

Chart 5: The lack of pitch.



When the data is split by school, there is a clear trend towards there being more likelihood of no pitch in poorer performing schools.

Chart 6: The link between 'no pitch' and school.



Finally, and related to the school performance, when looked at by the grades awarded, there is further indication that not making a pitch has some relationship to performance.

Chart 7: The link between 'no pitch' and performance.

It isn't clear whether the lack of pitch is, in itself, an indication of a weaker performance or whether it is more to do with the difficulty for the judge in fully understanding the quality of the learners' ideas if there is no pitch to listen to. But what is important is the need to know whether the lack of pitch is by accident, or on purpose – and in Phase 4 this problem will be addressed. What is also important is that learners are made aware of the value to them of making the pitch.

The special case of the primary school

In all of the above, one school needs particular mention – the primary school that took part in the trial. The learners involved were all either Year 5 or Year 6 – so considerably younger than their counterparts in the trial. While (as can be seen from the overview given in Chart 8) there was a certain incidence of what might be seen as an irrelevant pitch – typically talking about what was being done, not where the inspiration came from, it is impressive that all children made a pitch and that none said they didn't know where their ideas came from (2% of the total group did declare this). Whilst the handling collection leads their claims for inspiration, they have an interestingly similar, balanced approach to School 08 – the highest performing school – despite (not surprisingly) having one of the lowest mean performances.

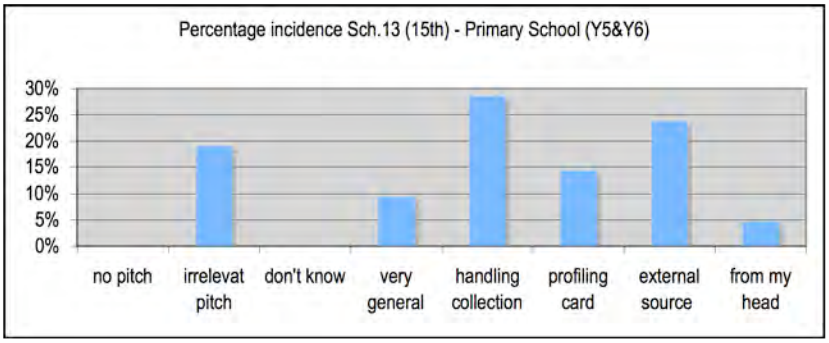


Chart 8: an overview of inspiration from the primary school.

Implications of the findings

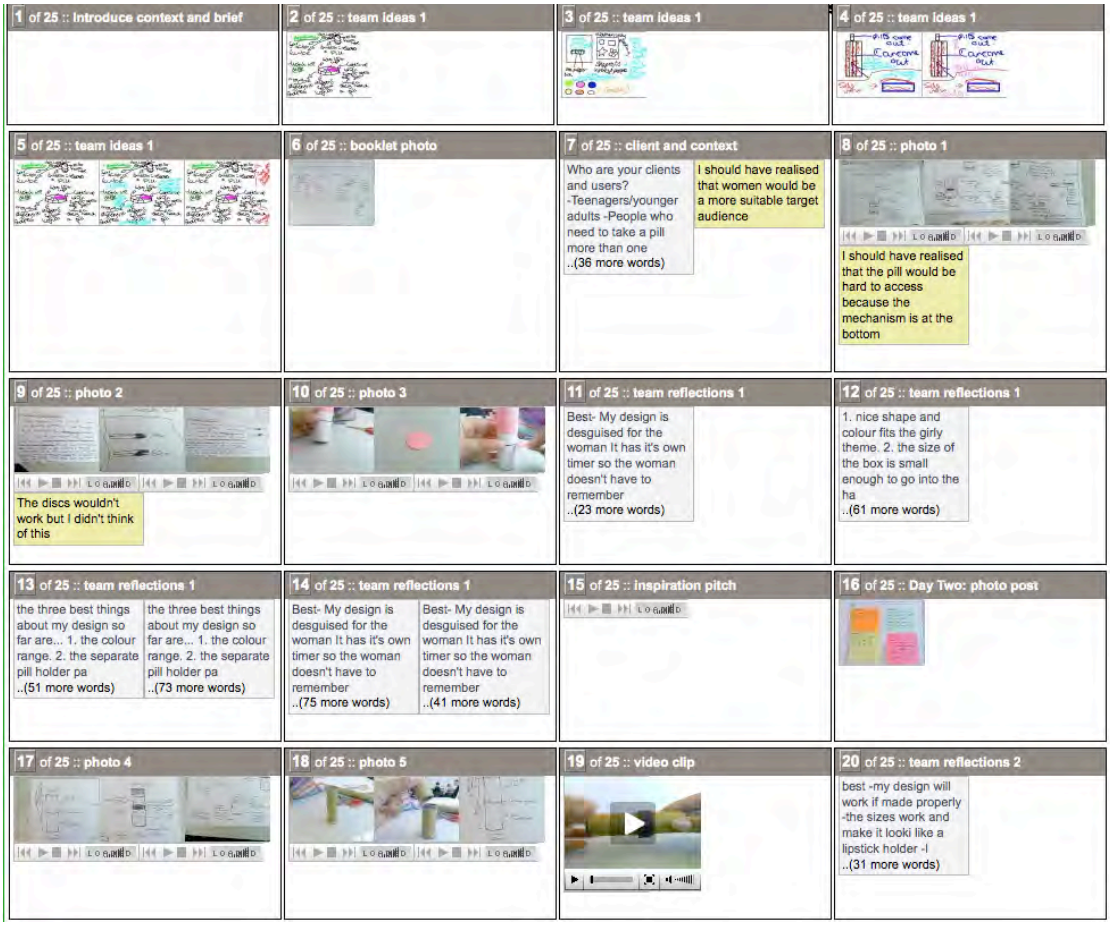
In summary, there are certain messages to be taken from the analysis of the inspiration pitch, relating equally to the future use of the approach for controlled assessment and more generally to learning and teaching in design and technology.

The need to ensure that learners have every chance to make a pitch, be confident it has been recorded and understand the consequences on not exploiting the opportunity the pitch presents to share their thinking and understanding.

- The need to strengthen the guidance given to teachers to ensure greater equity in the quality of the handling collections available to learners and the way they are introduced in the activity.
- To draw teacher's attention to the positive impact on performance a carefully selected handling collection can have.
- To build on the positive support afforded through the user details given through the profile cards – to continue their use in future activities and to indicate the value of referring to them – drawing to teachers' attention both the gender and the performance differences.

four case studies of complete portfolios

We now move to a more holistic presentation – essentially walking readers through four complete case-study portfolios. They have been chosen to represent competent to good performance and in the process to demonstrate the capability of the e-scape system to capture and exemplify performance in design & technology.



case study 1

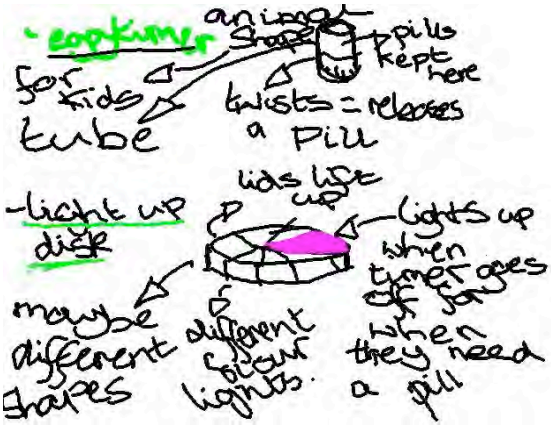
This student shows a particular strength in her understanding and application of the process of modelling an idea, using graphically sophisticated annotated sketches, alternating and interacting with quickly-made 3D prototypes that have enabled her to explore and resolve specific problems in the development of the product. Her thinking has clearly been joined-up and there is evidence of her awareness of the complexities and trade-offs needed. Although slightly weighted towards resolving technical issues there are some unusually perceptive insights into the user-experience, and some coverage of aesthetic issues.

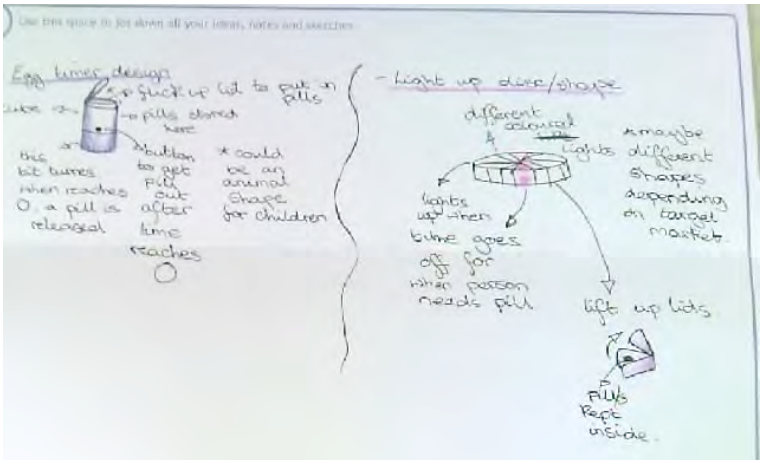
However some aspects of the design are perhaps inappropriate and unnecessary, and while much of the story of the development of her work is well expressed, some decisions and comments lack clarity. Generally she has played safe, sticking to her

initial idea and methodically converging towards a solution without considering more risky possibilities.

Analysing the process

Box 2 (First ideas on PDA
The student makes a lively start initially exploring 2 ideas in some detail, clearly thinking about the various product functions.





Box 6 (Booklet)
Here the student refines the ideas explored in Box 2, adding details, but does not move the basic concepts much further forward.

Box 7 (Client and context)
The student writes:

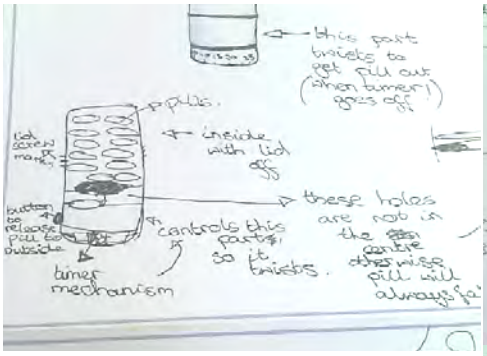
Who are your clients and users?
-Teenagers/younger adults
-People who need to take a pill more than one time in a day
-People who want to try and disguise the fact that they need to take a pill

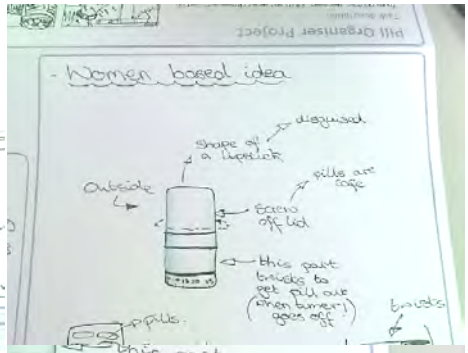
What will your product have to do to be successful?
-Be quite modern
-Not an obvious pill box
-Unique

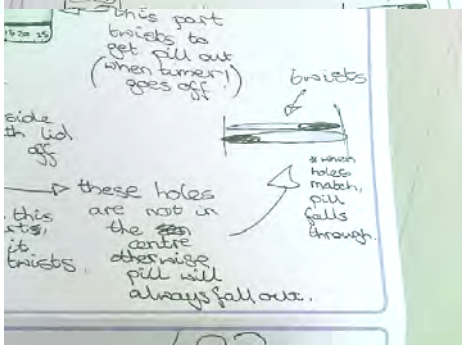
Rather than simply repeating one provided on the profile cards, the student defines her own market and the frequency of use of the product. Also added is the requirement for the device to be disguised in some way and for it to be 'quite modern' and 'unique'.

Box 8 (First developmental session)

Instead of launching into 3D, the student appropriately decides to spend more time developing ideas through sketching, extending the details of appearance, opening mechanism, dispensing and timing. The audio comments confirm that attention remains with developing the design.



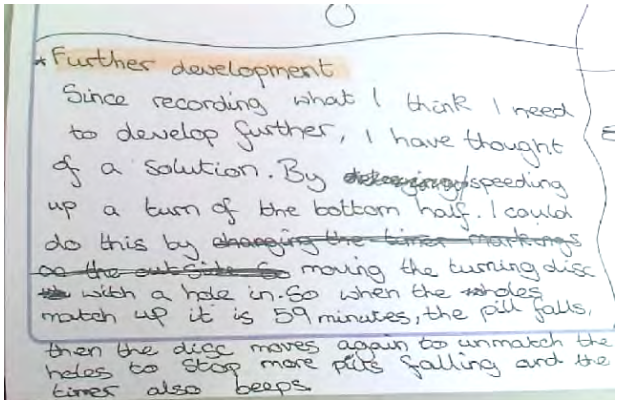




Audio response:
"I'm choosing my design as a lipstick holder so when they (ie women) need to take their pills they won't be embarrassed. I think I need to look at the mechanism as I'm concerned that more than one pill may come through when the timer is activated."

Box 9 (2nd developmental session)

Here the student records her thinking about the timer mechanism, and this is explained in more detail in the audio comments:



'This solution I've come up with I think will work for my design and help the client (user) to get the pill out.'

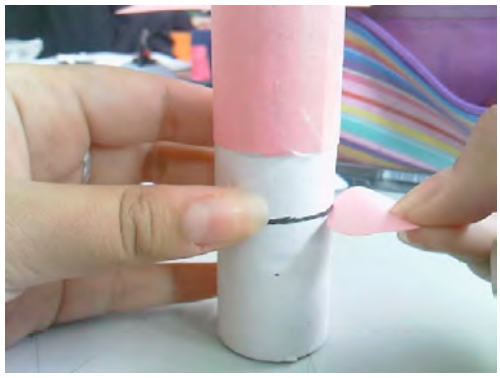
'Having come up with the solution I'm now concerned about the timings again because if I want the timer to be an hour long then the disk is going to have to move 3,600 times because there are 3,600 seconds in an hour.'

Box 10 (3rd developmental session)

Moving into 3D the student is clearly using the model effectively to explore and test how a possible mechanism might work, and the audio confirms that she is still focused on developing the design rather than making an appearance model.

'I am now pleased with my design as have decided to use a normal egg-timer mechanism and have changed the shape and form of my object to allow no (changes?) to the disk or mechanism.'

'I'm including in my design a piece to be pulled out to allow the pill to drop down. However a normal



clock mechanism takes up the whole of the base so I might not be able to access the pill from underneath so I need to develop a system or a way to get the pill out from the side of the design.'

Box 11 (Day 1 reflections)

"Best-

My design is designed for the woman

It has it's own timer so the woman doesn't have to remember when to take the pill

My design is innovative

Bad-

The opening for the pill may need to be improved

Making sure the pill drops"

Her text comments at the end of the first day confirm that she retains a tight grasp on what she is trying to achieve and has identified the priority for further development.

Box 15 Inspiration Pitch

'My design was inspired by an egg-timer and I'm incorporating this into my design by making a (it) whole compartment of my design.'

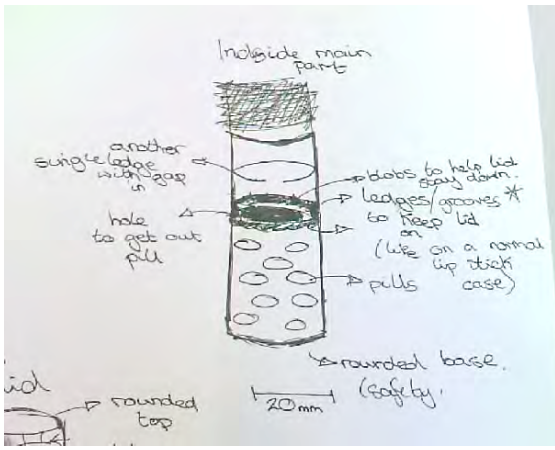
Here the student reveals the source of her initial inspiration and explains how it has been used in the design - though it is not made clear what aspects of the egg-timer she has adapted.

Box 17 (4th developmental session)

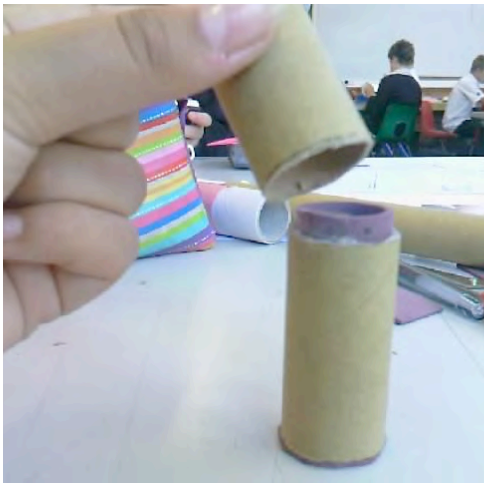
At the start of day 2 the student returns to sketching to explore new ideas for the timing and dispensing mechanisms. Her audio comments reveal her positive approach to her design development and confidence in her progress, and her recognition that it is time to move back into 3D modelling, with a specific developmental intention.

'I'm now really pleased with my design and I think it's going to work. I like my new idea of a block underneath the edge so that when the timer is twisting the lid is securely on and then when it's at zero the block and the hole are matched so that they can be released.'

'To further develop my idea I'm going to have to make a model so I can see if the sizes match.'



Box 18 (5th developmental session)



At this stage her audio continues to tell the story of her progress and success.

'And here's my model. I've worked out that the sizes do actually match, and it's looking pretty good.

I need to finish off the model and maybe extend the top half of it so it fits better.'

Box 19 (end of 6th developmental session)

The prototype is sequentially demonstrated, with the following audio explanation:

'Here's my final product. As you can see it's based on the size of a lipstick holder. There's a tiny hole here that can be twisted for people trying to get their pill. If I open it and take the pill out, there's a ledge with in a gap so this fits here to go through at the top. And then twist and put in place. At the top here there's a clock mechanism that turns the top bit so you can get the pill out.'

Box 20 (end of day 2 reflections)

"best

-my design will work if made properly

-the sizes work and make it look like a lipstick holder

-I really like the block and ledge idea to keep the lid on etc

need development

-even more rounded edges

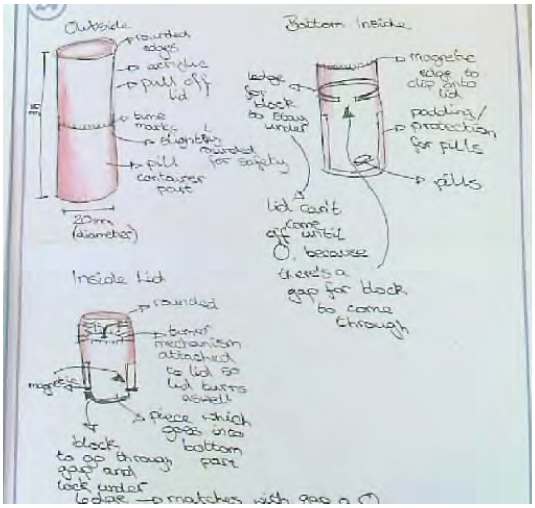
-thinner block to make it easier to put lid on

-colour in more"

These comments indicate an awareness of the need for further development in terms of ease-of-use and aesthetic appeal.

Box 24 (Fast Forward)

Fast forward sketches effectively start to develop the appearance of the product and its shape. She continues to develop the mechanism and adds refinements such as padding for the pills. The clearly annotated sketches are graphically strong, showing cut-away drawings and subtly indicating colour and form.



Box 25 (Lift Pitch)
"This is the opportunity of a lifetime. It's a pill-dispenser aimed at women who may be embarrassed about taking pills more than once a day. So I've designed it as a lipstick holder they can carry round their pills wherever they go. Also using the timer a bell will go off when the desired time is chosen. I'm using a unique locking system so the pills can never fall out."

The lift pitch appropriately highlights the market opportunity and the key design features.

Final review:
The student comments on 3 boxes, identifying key moments when her thinking went in the wrong direction.

7: I should have realised that women would be a more suitable target audience

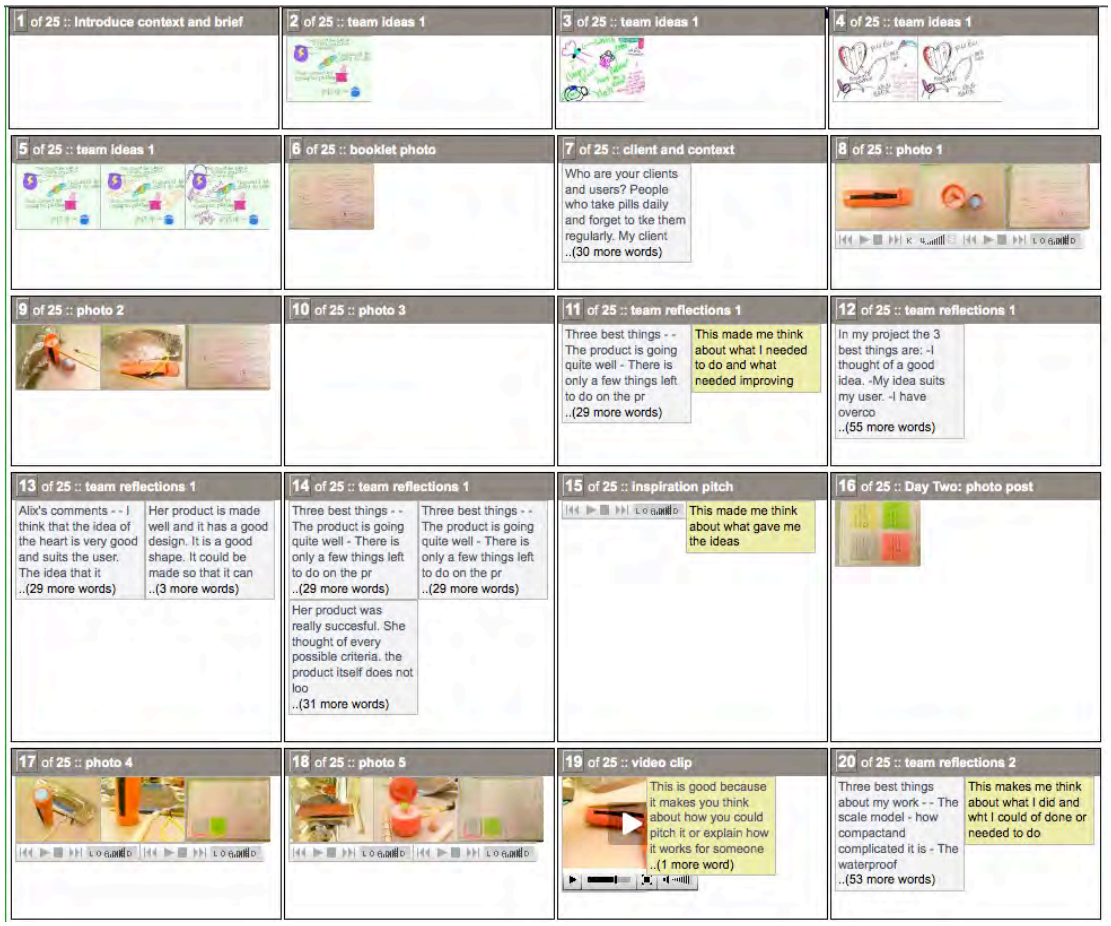
8: I should have realised that the pill would be hard to access because the mechanism is at the bottom

9: The discs wouldn't work but I didn't think of this.

Assessing the product
In terms of responding to the initial brief the requirements for the device to contain and carry pills have been well met as the product is clearly portable and the pills are securely retained. The operations of filling and dispensing have been taken into consideration, but the product has not so far been designed to make it particularly easy to do so. The solution is very good in terms of creating a form that is highly desirable, responding well to the issue of potential user 'embarrassment', though the solution does potentially come across as being more of a 'novelty' item.

A lot of attention has been paid to the development of the timing mechanism to release pill - but this is not really necessary, given that it is not too difficult to open by hand when alerted to, and has added unnecessary complexity. An alarm alerts the user after an hour, but there is no facility for alternative timings needed for different types and frequencies of pills.

The product would be distinctive and original in the marketplace, and is realistically manufacturable, though would probably need to be a 'high-value' product. It might make a good promotional product for a cosmetics counter.

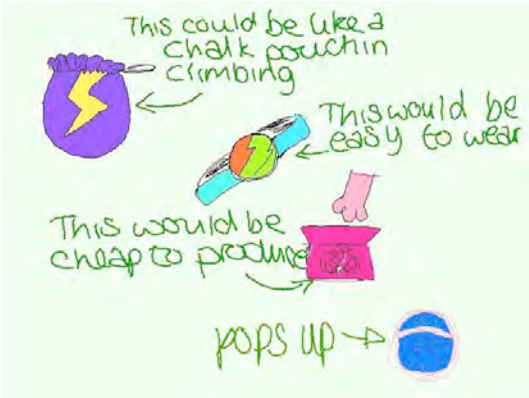


case study 2

This student begins with a range of ideas that she evaluates and then chooses one to develop further. In doing so she unusually explores more than one item, i.e.. the main container, a hooking/carrying device and a protective waterproof pouch. The Fast forward response is of particular note, extending her thinking into graphics and packaging. Her journey comes across as focused, organised and confident, though there are some gaps in the story of her decision-making.

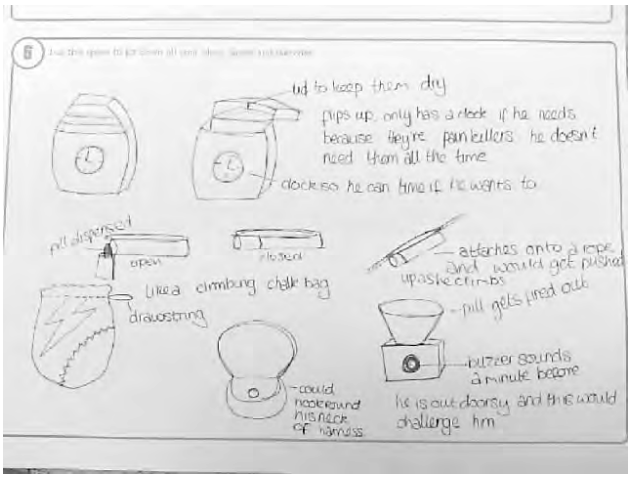
Analysing the process

Box 2 (First ideas)
The student begins with four ideas, none of which are particularly 'wacky', but the annotations indicate she has been thinking about carrying, containing and manufacturing costs.



Box 6 (Initial development on paper)

This develops well from the earlier box 2 drawings, rejecting one idea, considering two in more detail and adding two further ideas. The controlled drawings and annotations show initial coverage of issues of timing and the possible need for an alarm, dispensing and carrying.



Box 7 (Defining the clients)

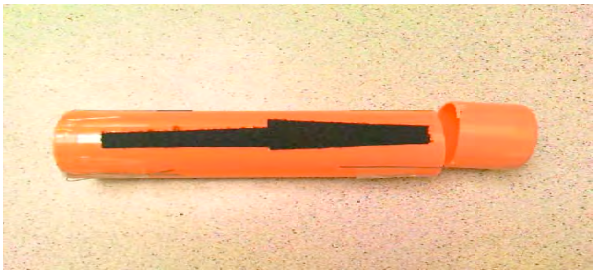
Who are your clients and users?

People who take pills daily and forget to take them regularly. My client is Stuart who does a lot of sport.

What will your product have to do to be successful?

Able to be used by people who do a lot of sports

Box 8 (Initial development)



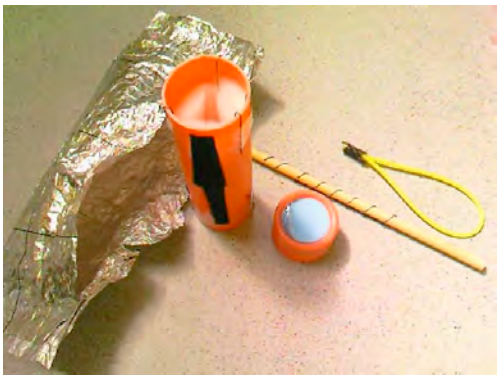
Audio response:

"What's working well and why - I think that the fabric inside my tube is working well because it's easy to put in and will dry up all the moisture."

"What needs further development - I think that the lid needs further development because the foam inside it has been dissolved by the glue."

The student moves to modeling in 3D and considers the issue of how to make the container waterproof. Unfortunately there is no documentation of why this idea was chosen or the approach to modelling adopted.

Box 9 (Further development)



The audio is missing at this point, but the photos reveal that rather than concentrating on the container, she has started to explore the way in which the container will be carried, and protected.

Box 11 (Reflection at the end of day 1)

Three best things -

- The product is going quite well

- There is only a few things left to do on the prototype

- The camera pictures are very clear and can be used to draw on and note on.

Box 15 (Inspiration)

"What inspired me and how I might try and make it work - I got inspired by the (profile) card that was given to me and the guy on the card said he liked outdoorsy stuff so I included all of the stuff which he said that he did and things which would help him carry the dispenser around with him and prevent it from being broken or damaged in any way."

This documents the inspiration for the choice of user, but does not explain where the ideas for her different designs came from.

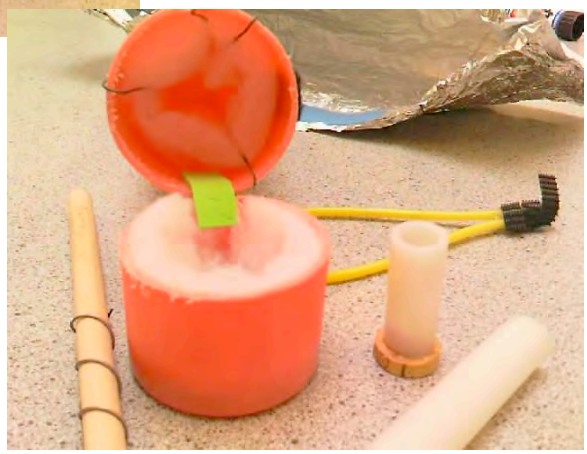
Box 17 (Day two development)

Audio response:

"What is working well and why - I think that my scale model and not my working model is going quite well because it looks very good.

What needs further development - I think that the scale model needs further development as it is only a tube that is cut in half at the moment."

The audio reveals that the student has started to make an actual size scale model.



Box 18 (Further development)

Audio response:

"What is working well and why - I think that drilling into my lid of my scale model is working quite well. Except that I have to find a better sized drill-bit because the drill-bit is too small to stick the wire in to make a hinge mechanism. The fabric inside the tube is working quite well to dry up all the stuff from the specification of the client."
"What needs further development - The shoe needs further development as the hinge needs to open that and I need to drill a bigger hole into the scale model to make the wire fit into it so that it has a hinge mechanism. The cork in the end needs to be tightened and glued in."

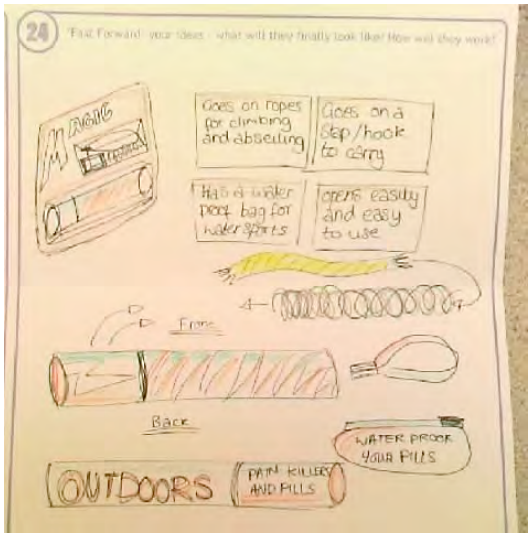
Box 19 (Video presentation)
"This is my scale model which opens and closes and the pills go inside. Then there is a model that has some fabric inside it to help keep the pills dry. There is velcro on the top so it can go onto the spring that can slide up and down a rope or a hook to attach onto a bicycle or something. And there's a waterproof bag that keeps it and it does up like this. Et voila!"

This presentation is very focused and the student is clearly very articulate in describing the features of her design and expressing her confidence in it as a solution. However, exactly how the rope-moving system (as identified in box 6) works is still not clear.

Box 20 (End of day 2 review)
Three best things about my work -
- The scale model - how compact and complicated it is
- The waterproof bag which is made of thermal insulation
- The hoop handle which makes it different from other designs

Three things which could be improved -
- The bag could be more stylised because it is a bit rough around the edges
- The rope moving system which could be improved
- There could be more add ons

Box 24 (Fast forward)



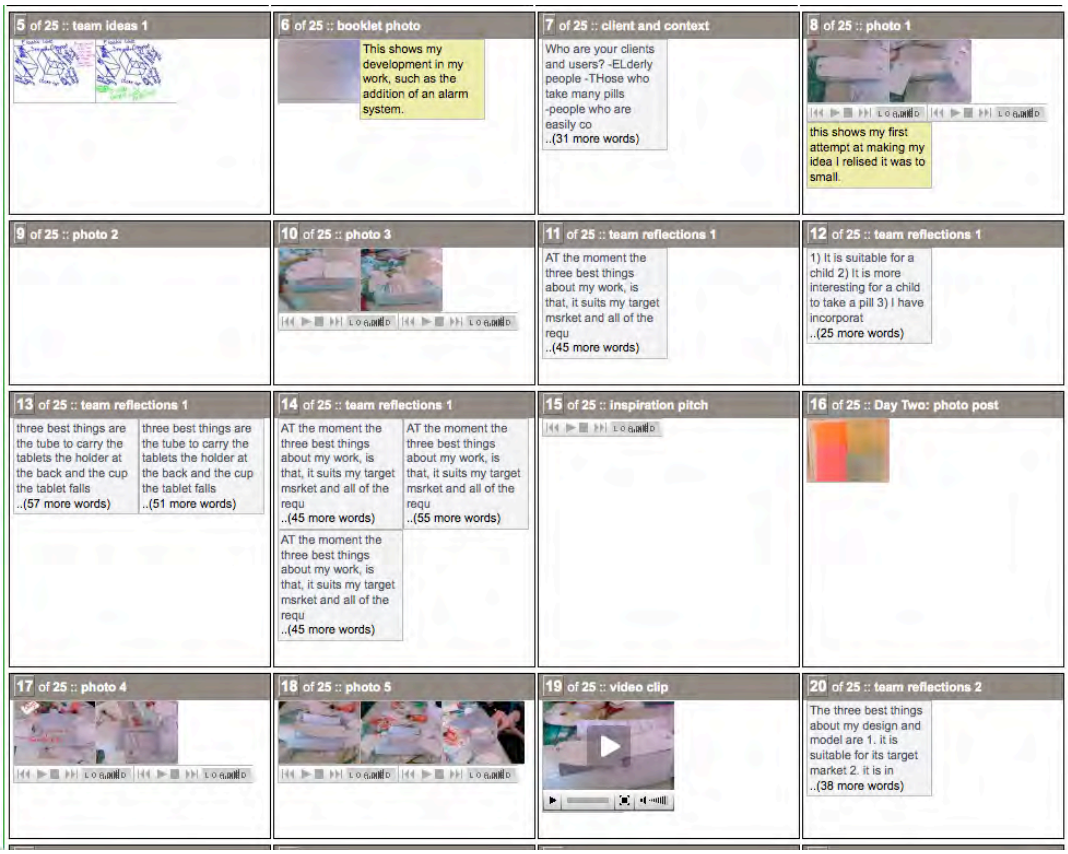
This response effectively moves the idea further forward. Rather than developing the shape and form of the container it instead considers its presentation in the market in terms of packaging, graphics and its features and benefits, which in turn follow through the aspects of the design she has been focused on - how its carried, kept waterproof and is easy to use.

Box 25 (Lift pitch)

"Hi - I want you to view my product as a breakthrough in pill dispensers and it's very good it comes with a waterproof bag, hoop and rope-climbing equipment. Here are some pictures."

Assessing the product

Aspects of carrying and containing the pills have been given particular attention, and the product appears highly appropriate for the 'active sports' market it is intended for.



case-study 3

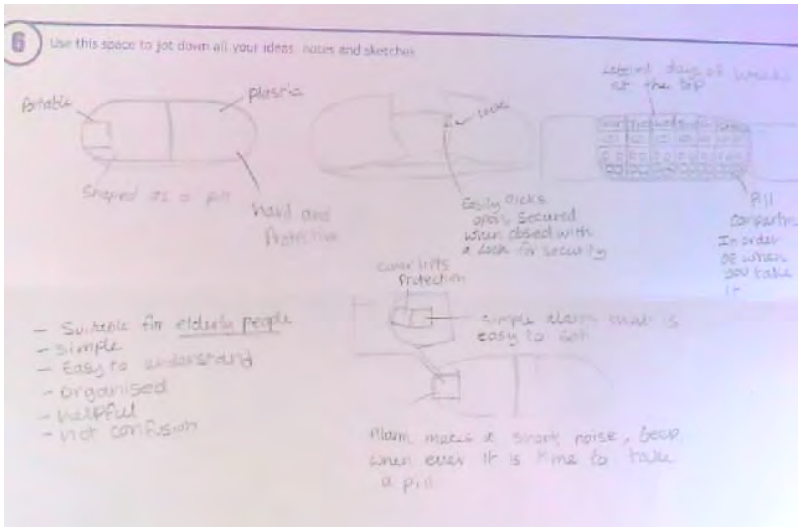
This portfolio is striking in two particular respects. First is the student's desire to fulfill the needs of the user; and this drives the development. Second is the clear grasp of the use of sketches and rough and ready prototypes - or 'soft' models - to enable the construction of quick and easy 3D representations. There is clear growth from the first sketches on the pda through to the fast-forward drawings, and a sense of pace and sustained application. Some opportunities have been missed, for example in terms of the colour and texture of the product, and to appropriately 'sell' the product in the lift pitch.

assessing the process

Box 2 (Initial ideas on PDA)

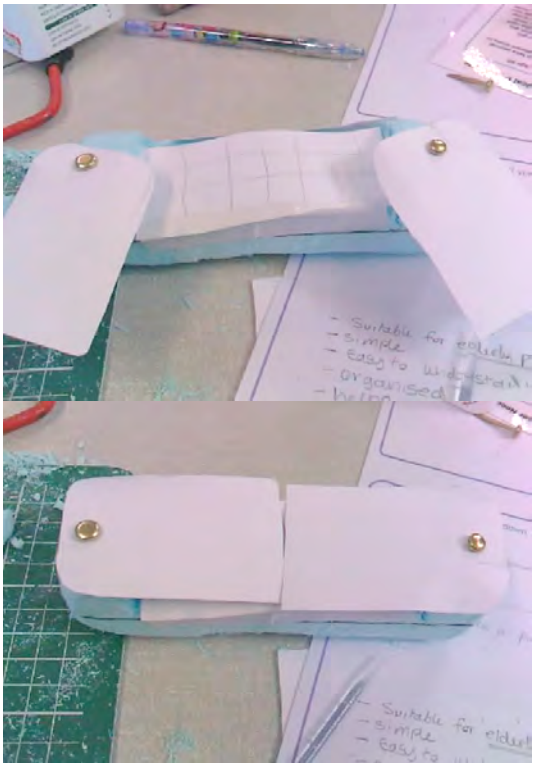
Initially the student goes for a fairly predictable 'box with compartments' approach, but the annotation identifies an unusually wide range of issues that are already being considered - materials, size, appearance, closing mechanism, labelling for identification.

Box 6 (Developing ideas on paper)



The immediate development on paper is to adopt the shape of a pill as the form of the container. The drawings use different and appropriate graphic conventions to model different aspects of the design. The annotations continue to reveal the breadth of thinking going on - the characteristics of the material, the closing mechanism "easily clicks open, secured when closed with a lock for security", and the addition of a 'simple alarm that is easy to set', explicitly related to the self-defined needs of the user - "elderly people" - to be easy to understand, and organised.

Box 8 (First 3D development session)



Audio:

"Mine is working well because its practical and it serves its purpose, to hold lots of pills and to also be simple and easy to follow. It's also it's quite compatible and portable and it opens easily and its secure because it has a lock. Also because it's in different sections and it has the different days on it, it will be really easy to follow for someone who is an elderly person and that's the market I want to focus on."

"The only thing which needs firm development on my design is that it needs to be completed and it needs two side panels on the side so no pills fall out of it and it will be a lot more secure that way and also I need to add some colour just to make it a bit more interesting to look at."

The photographs reveal that this is clearly a 'rough and ready' developmental model, made to explore the shape, form and proportions of the container. As such it is made from card, paper and paper-clips - materials and components that are quick and easy to work with. The audio reveals the student's awareness of the lack of visual appeal.

Box 10 (Further development)

In this development session the student has focused on adding the compartments to hold the pills for each day of the week.

Audio:

"I've added bits of cardboard to separate them into different days - and I've started to write the days on each section so it's more organised."

Box 11 (end of Day 1 review)

"At the moment the three best things about my work, is that, it suits my target market and all of the requirements elderly people desired in a pill pot. For example it is practical and portable, and also is well organised so the person would not get confused. As for improvement it is not very attractive, and needs colour and neatening, also the days enlisted."

The review reinforces the firm grip the student has on meeting the requirements of the user, and the specific aspects that need further development.

Box 12 (Inspiration)

"I got my idea from the laminated piece of card saying the pills an old lady takes each day and how confusing it is to take them all and how she like doesn't know when or what pills to take at sometimes. So I thought maybe to make something which was organised which could be easy to follow would be a lot helpful and would, yeah, fulfill the needs of somebody of that age."

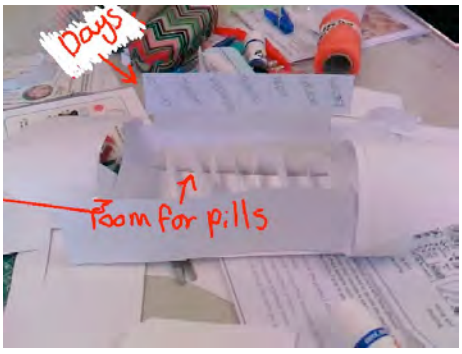
The emphasis on understanding and being motivated by the intention to fulfill the specific needs of the user is commendable.

Box 17/18 (Day Two development)

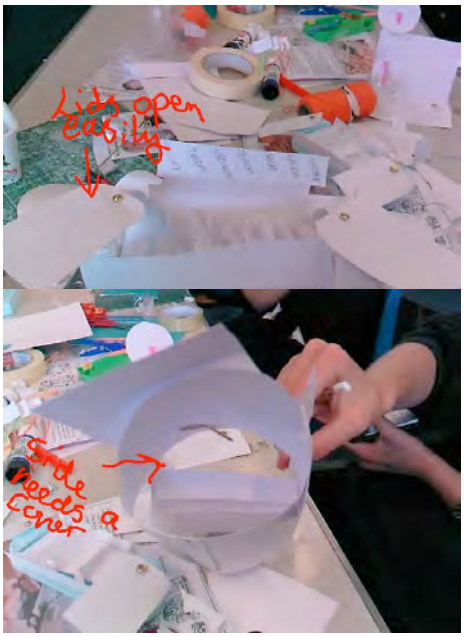
Audio:

"It's working well because I started again with a different material which is a lot more flexible and it looks a lot neater and also it's larger so you can fit different sized pills in it, which is quite a good quality considering the age range which I'm working for.

"For further development I just need to add a lid so it's all covered and protected and some colour."



Box 18 (Further development



"I have added a lid which easily slides open so that avoids any pills falling out of the pot."

"On the sides I need to add little cover-ups so it looks a bit more presentable. Yeah."

At the start of day 2 a second model is constructed, appropriately changing the material and the scale. This causes a significant development in the shape and form of the dispenser as the ends become vertically circular. The alarm and lid-locking mechanism are recognised as being standard components are therefore sensibly 'black-boxed'.

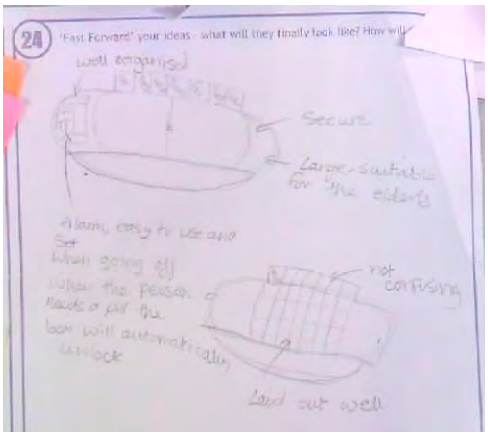
Note how the student very effectively uses the technical facility of the pda to add annotations to the photo to help explain her design. The repetition of 'yeah' is a positive indication that she feels pleased that her design is successful.

Box 19 (Video)
"This is my pill container. It's designed for an old person so it's quite easy to use. For example inside it has sections underneath each day so that they can follow that well. There is also an alarm system which is easy to use and easy to control. It's enlarged so that also makes its easier to use and...is practical for someone of that age."

The video shows the product being turned to show the various features being described. It is both informative - packing a great deal into a short space of time - compelling and convincing.

Box 20 (Reflection)
The three best things about my design and model are
1. it is suitable for its target market
2. it is interesting to look at
3. it is organised and simple to understand

The three things I need to improve on are
1. the portability of the case
2. if the pills will be protected
3. the lack of simpleness in its design



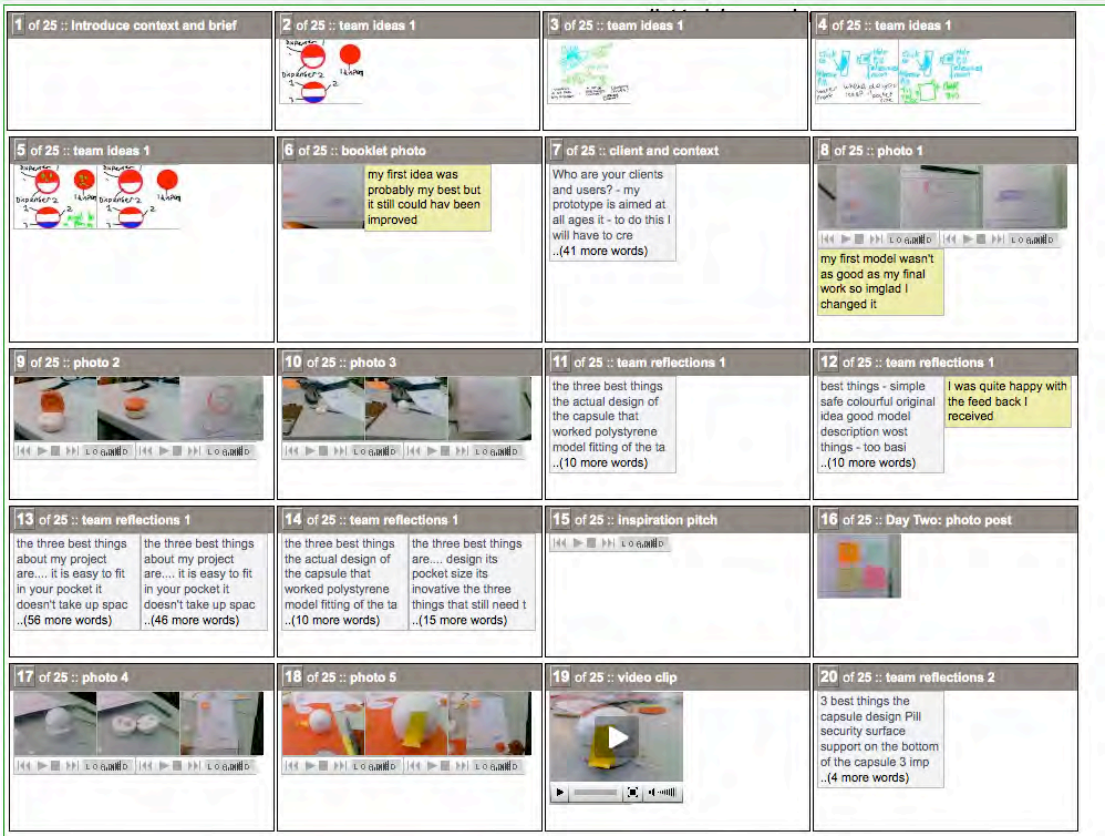
Box 24 (Fast Forward)
In terms of features, little is added at this stage. Instead, however, the drawings begin to give a sense of how the development of the model into its appearance as a manufactured product
Box 25 (Lift pitch)

"My pill container is good because it is beneficial to elderly people who have to take lots of medication. It's nicer."

The lift pitch is disappointing and failing to suggest how it might be commercially successful.

Assessing the product

The design successfully tackles the problems of containing and organising the pills in a simple way, meeting the requirements of the brief well. The recognition and understanding of the needs of an elderly person are commendable. The final drawings suggest the development of a real product that could be manufactured and sold.



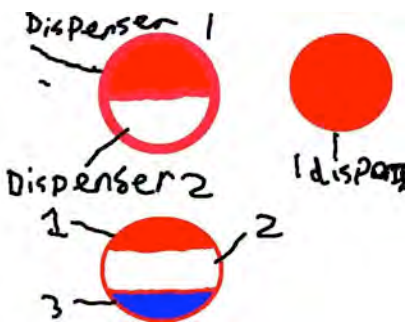
case study 4

This student displays an unusual awareness of his preferred method of modeling through drawing, and his audio comments reveal how development has been driven by the need to address design issues rather than simply making a presentation model. He sees complexity and there is evidence of optimisation and a clear sense of joined-up thought and action and of the ability to identify emerging strengths and weaknesses. Although not particularly creative, the product is convincingly realistic and not over-complex.

analysing the process

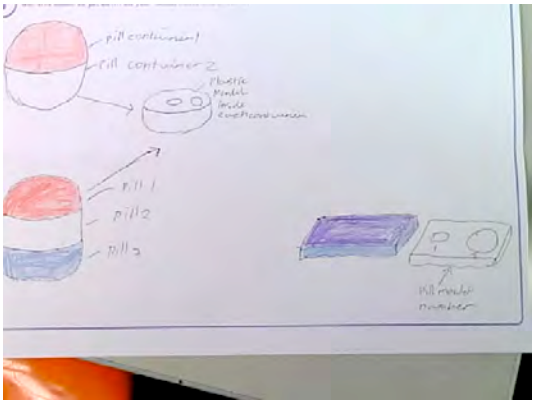
Box 2 (First ideas on PDA)

The first drawing does not reveal a great deal about the thinking at this point, except that it's circular and has a number of compartments.



Box 6 (Ideas on paper)

These sketches reveal a great deal more - actually the intended shape is spherical with the pills contained in moulded plastic inserts in the top, bottom, and possible middle sections.



Box 7 (Clients and users)

Who are your clients and users?

- my prototype is aimed at all ages it

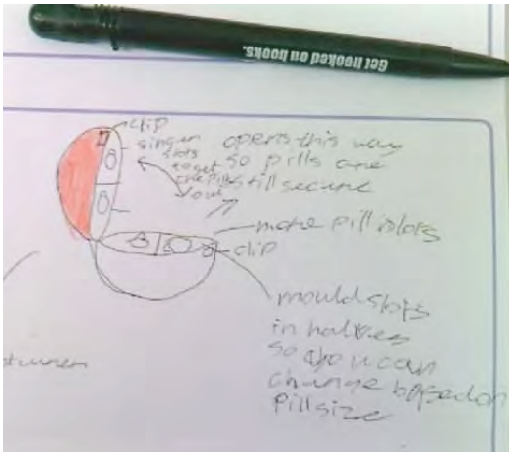
- to do this I will have to create models which allows different amounts of pills

What will your product have to do to be successful?

it will have to appeal to all ages maybe with different colours and labels for get confused with the order in which they take.

Here the student moves away from the profile card prompts, extending uses to all ages. The idea that it should be 'modular' to allow for different amounts of pills is clarified, as is the need for a system of colour-coding and labelling.

Box 8 (First ideas on paper)



Instead of moving immediately to 3D modelling, the student now develops his idea further on paper, thinking about the need for a clip for the two halves the 'secure-ness' of the pills, and adaptations for pill-sizes.

Audio response:

"Photo 1 seems to be working with the two ideas there - I've redeveloped that into where the pills are going to go - how I'm going to open the casing and that - where I'm going to store all the pills and that - and yeah that's the redevelopment in photo 1 to photo 2. Photo 3 I have decided that it would be probably my main idea .

Photo 1 still needs a little more development because it does look like it's going to be a complicated model to make...

Moulding - getting the right moulds for each pill's going to be difficult.
But - apart from that it is going to be a bit of a challenge to get that sorted."

The audio reveals that he is thinking ahead and realising that making the model is not going to be easy, but that he's willing to accept the challenge.

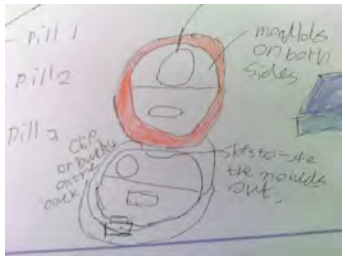
Box 9 (Further development)

Audio response:

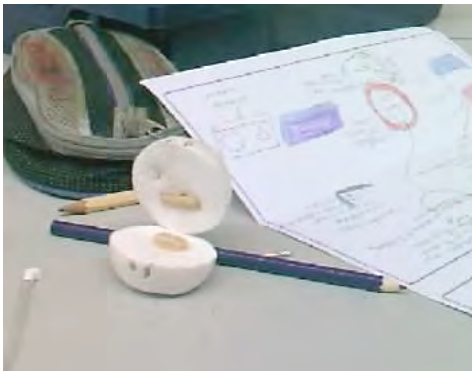
The actual design - the actual plan of it is working up to standard. That's going to be the easiest part done That's done. The model still needs work on it but it's coming along alright. The reason the plans work is that I can work with drawings and I can associate what bits need to go where with drawings and that and it's simple enough.

The clay model that I made needs further development. It's not working too well, but when I get to the actual prototype I will be creating it with polystyrene as it looks a bit like a toilet seat at the moment!

The student's analysis and awareness of his own capabilities in terms of developing ideas through drawings is of particular note. He is also planning ahead in terms of recognising the inadequacies of his first model and the need to construct a different one using a different material.



Box 10 (Further development) By the end of this session the second model has been constructed - considerable progress has been made during the first day, and the student is aware of the issues that still need to be resolved.



Audio response:

"A model I've replaced seems to be working well with polystyrene now instead of clay - clay was a bit difficult to control and shape it and everything.

Now that I've switched to polystyrene it's looking good. I've put two wooden
tablet replicas in - I don't know what you call them but that's what I'm going to call them and they fit in very well. I'm going to get the other slots sorted in so that they fit more tablets in basically - yes that is about it.

What needs further development? Well basically the model again needs development because it has a few bits and pieces that still need sorting out such as maybe more design - more tablet space, etc, etc.

Closing it properly will need to be an issue here - getting a layer on top - coloured in or something - maybe put paper over it so it's covered."

Box 11 (Day 1 reflections)

the three best things

the actual design of the capsule that worked

polystyrene model

fitting of the tablets

the three improvements

the model shape

open close mechanism

colour of the model

Box 15 (Inspiration)

"The thing that inspired my design was one of those twisty air freshener things. Instead of twisting it I thought I'd flip it over, so that - it looks like - it's shaped like a capsule. And what I wanted to do with it was make it easy to use easy to handle and easy to take tables out of basically."

As well as identifying the source of the shape and form of the container, the student explains how he adapted the concept to provide a starting point for meeting the user-needs for a pill container.

Box 17 (Day 2 development)

Audio:
"*The new model that I've created is working well and where the pills go is working well - I've sorted that out - so no need to worry about that any more. That's basically it.*

I think that how the capsule closes and how it's actually attached to you will need some more development. But apart from that it's working well."

Box 18 (Further development)

Audio:

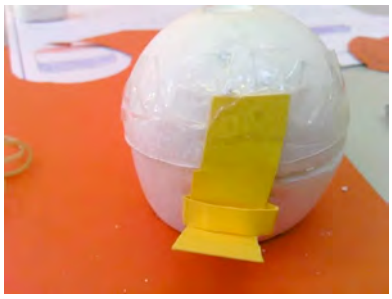
"The model is coming along nicely with the new lock I've put in so that it stays secure and all the pills secure. That's one thing that's working well.

The whole design yeah - it does need to have development. Some parts are a bit damaged from attempts of creating the new lock and that - yeah it still needs development.

I've scrubbed the idea of the moulds for the pills because I've created a space for each size of pill - because I believe it is a good idea and I will continue with. Maybe add a bit of colour to it - maybe sort out the open and close contraption."

Box 19 Video presentation

Audio:



"Here's the lock that keeps the tablets secure. Here are the tablets with finger holes so that they can be taken out. There's a label for the tablets and here's a support for when you need to stick the labels down. My innovative design is very, very easy to use and it is pocket-sized and you could put it in your bag as well."

Box 20 Final review

3 best things

the capsule design

Pill security

surface support on the bottom of the capsule

3 improvements

the lock

open and close

labels

Box 24 Fast forward

Fast-forward reveals some further detailing, including further flattening the shape at the bottom, changing the location of the label, and adding a belt attachment.

B**ox 25 Lift pitch**

"Alright Well - Here's what I am going to be presenting to you today. This is a capsule which can carry up to 4 pills of all different sizes. It's innovative, creative and it has many different colours and designs. You can buy these at a very reasonable price and it is very - very - very creative. It saves time and is pocket-sized and you can put it in your bag as well and take it around with you."

Here the student adopts an appropriate register for the pitch, highlighting some the key features and pitching its innovative qualities alongside its economic viability.

Assessing the product

The student has considered all the major requirements covered in the brief - containing, carrying, filling and dispensing, and appearance, and the general need for 'ease of use' permeates the development. In aiming for a wider target market the need for an alarm and timing system has been removed. The product is convincing as being potentially marketable.

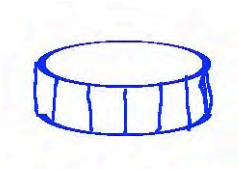
characterising performance levels box by box

In the previous section we looked in detail at four complete case studies of performance as captured in four portfolios. In this section we take a different look at **the whole sample** of portfolios. We use this wide-ranging analysis to examine the general performance of the whole sample of learners a box at a time, working through a number of portfolios. In each case we analyse the performance in *good* portfolios, in *poor* portfolios and some in the *mid* range. The purpose of this section is to begin to explain the interdependence of each component of individual portfolios and how these combine in radically different ways to build overall capability. Just because a learner responds well to the individual tasks does not necessarily mean they can join these up to form a good overall picture of design ability.

Box 2 (First ideas)

Poor performance

The first example shows a single, minimal drawing with no annotation, but indicating some sort of circular container with compartments.



The second example uses a crude plan view to represent some sort of container with compartments for morning and afternoon.

mid / poor performance

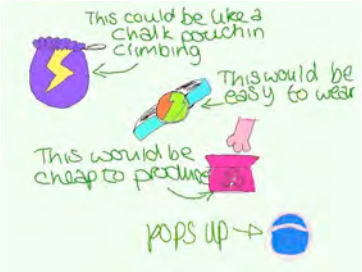
The student on the left produces one idea that shows a very obvious and unimaginative solution, but nonetheless clearly communicates the intention using a 3D sketch and basic annotation.

This student also produces a single idea, very closely based on an existing solution. However the use of a 3D sketch and a 'close-up' side elevation, and the arrow showing the direction of movement suggest some knowledge of graphic communication

techniques, and the annotations indicate some depth of thought.

mid / good performance

Unusually the student on the right conveys an immediate concern with form and colour, managing to produce a bold 3D representation using the pda graphics program.



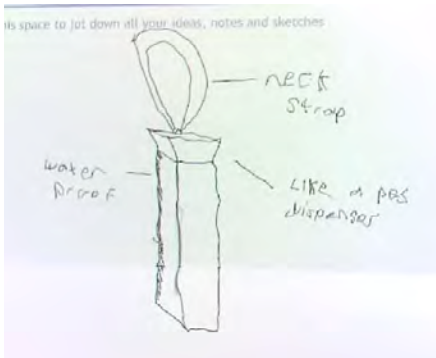
Meanwhile the student on the left shows four different but fairly conventional ideas, with the annotation revealing that she has started to consider the key design issues.

very good performance

The student above shows excellent first ideas, exploring structural and visual concepts derived from nature, while the one on the right explores one whacky idea in good detail - including how a complex mechanism might actually work. This is remarkable considering the time allowed and the limitations of pda drawing program.

Box 6 (Booklet ideas)

Poor performance

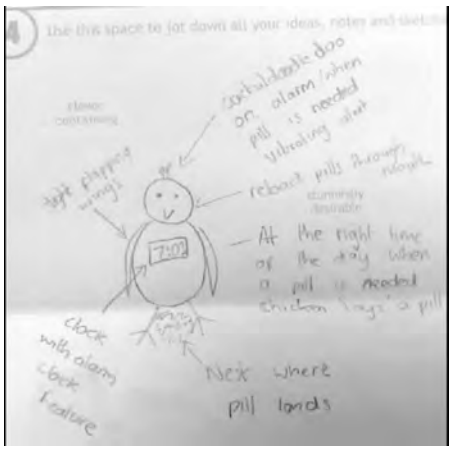


This drawing is essentially a repeat of the representation in Box 2, adding virtually nothing to the concept. It does however indicate a neck strap and the fact that it is waterproof, showing that the student has thought about aspects of carrying and protection, with filling and dispensing being achieved using the same mechanism as a Pez dispenser.

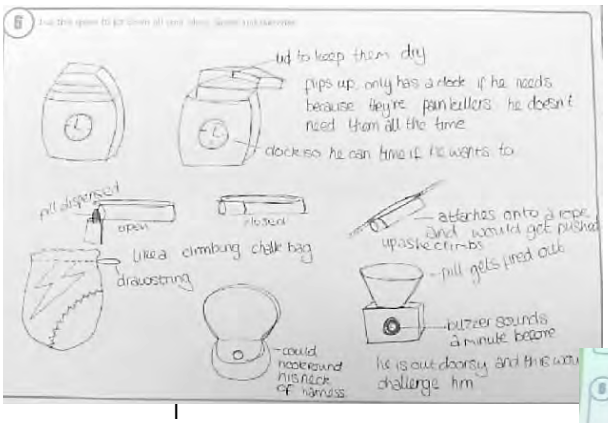
The student has also taken the initial box 2 ideas and essentially re-drawn them on paper. However the annotation reveals some limited consideration in terms of filling and dispensing and visual appeal.

Mid level performance

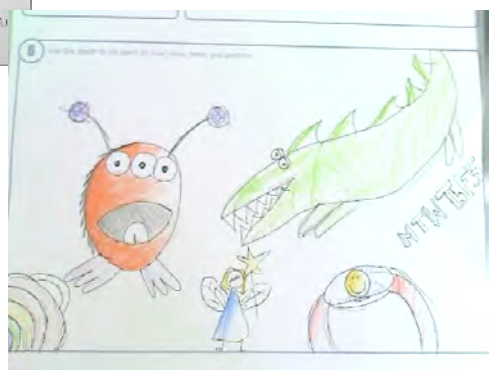
One idea, shows some imagination and some risk-taking but pursuing an unconventional approach, with an emphasis on attractiveness for a young child. The annotations reveal considerations of what needs to work, but not how those things might be achieved.



mid-good performance

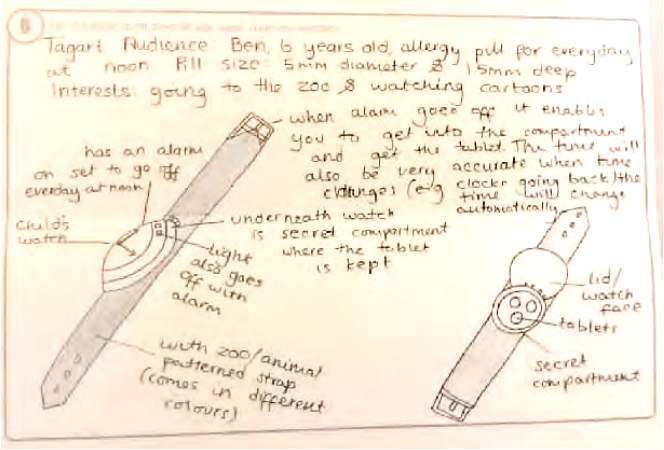


This develops well from the earlier box 2 drawings, rejecting one idea, considering two in more detail and adding two further ideas. The controlled drawings and annotations show initial coverage of issues of timing and the possible need for an alarm, dispensing and carrying.



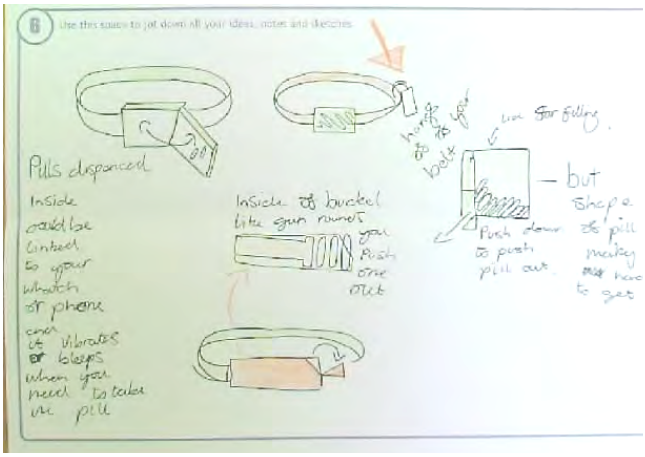
though this drawing includes no annotations, the ideas are obviously highly imaginative as starting points for forms that will be attractive to young children. Subsequent annotation reveals the range of wider issues being considered as these initial sketches are being externalised.

Very good performance



A good mixture of text and controlled drawings used to add detail to the initial idea. The brief is re-written, and two drawings - one a plan view and the other a side elevation - cover the need for a visual and audio alarm, an attractive strap and a 'secret' compartment. Colour is used sparingly but effectively.

These sketches are flowing across the sheet as ideas are being rapidly explored and developed. Thoughts about linking the dispensing mechanism to a watch or phone alarm system move to cross-sectional sketches that focus on a dispensing mechanism 'like a gun round' that would push each pill out. Arrows very effectively help describe movement.



Box 7 (Client and Users)

This box is essentially for focusing student's minds on a particular client group and their needs. However it also provides an indication of their ability to think further, beyond the stated requirements in the brief and on the profile cards.

Poor performance

Who are your clients and users?
-Pill users
-Children

What will your product have to do to be successful?
-carry pills in a sufficient way
-look attractive to kids

Here the student provides a very superficial statement of the intended clients and their needs.

Mid performance

Who are your clients and users?
My intended users are people who suffer migraines and need to take 2 pills - 1 first and another 4 hours after.

What will your product have to do to be successful?

Tell the user when to take the second pill and to easily dispense the pills and contain at least 5 of each pill.

The clients and users are repeated from the profile card. The brief is extended by the need for a reminder system of some sort, and that there needs to be provision for five sets of pills, though no explanation is given for this number.

Mid – good performance

Who are your clients and users?
- it would be any extreme sports men and women that usually take pain killers
- maybe business men or women that wear a belt and also need to take pills regularly

What will your product have to do to be successful?
- be compact so its not big and in the way of everyday things
- it would have to be easy to use and easy to fill up again
- it would also need to withstand any damage if it fell on the floor.

Here the requirements for success have been more fully specified, largely drawing on the original brief, but adding compactness and the need for it to be able to withstand damage.

Who are your clients and users?
- child aged 6
- likes animals
- scared of losing pill

What will your product have to do to be successful?
- animal theme
- secure so wont loose pill
- ergonomical smooth, no sharp edges.

The student has added two performance criteria not mentioned in the brief or profile cards - that a young child might be anxious about losing their medication, and that the container will need to have handling properties that will not include sharp edges / relate to the user.

Very good performance

Who are your clients and users?
- the clients may be a company that is interested in the product
- parents of the children who the product is aimed at
- the user is the child who has need of this product and who would like to use it.

And:

Who are your clients and users?
- Parents would buy the product for their children.
- Adults would buy the product for themselves
- Companies and individual care homes and hospitals would buy for older residents.

These students clearly understand the difference between clients and users, and the latter extends the range of the potential market.

Boxes 8/9/10 (Day 1 Modelling development)

These sessions produce a great variety of approaches that cannot all be exemplified here. While most students decide to move directly to 3D modelling, others continue to develop their ideas on paper. The majority proceed to make one product, while others might make several. In themselves, these different approaches do not have a significant impact on the overall quality of development and the evidence of their understanding of the process of developing a design idea through modelling.

Poor performance

"My further development is to put all the sides of my box together and glue them all up

"I've got to start again as it all went wrong, so I'm starting from scratch"

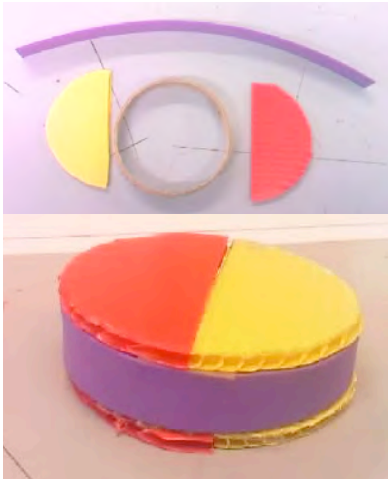
During the morning the student has struggled just to cut out the sides of a square box she is trying to make.

"The look of the model is looking good because it's all come together properly."

"I need to do the lid of the model."

The student has moved from their initial box 6 drawing straight into the manufacture of a 'presentation' model, with no further design development. The audio is minimal, revealing very little analysis of process. After some initial enthusiasm in making the basic structure, the pace slows during the morning with increasingly less progress being made.

mid performance



"I've worked out how to make my model and what I'm using and I've got all my pieces and I've just got to work out how to put them together."

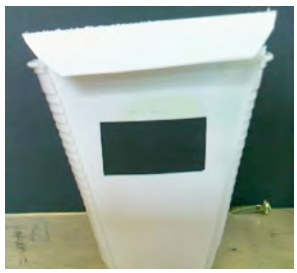
"The thing that needs further development is my whole thing. I need to put it together. I need to finish the base of my product and then test it."

"What's working well is the shell of the product it's looking good."

"The further development needed is the inside of it and how I actually get the pills out of it".

This student makes fair progress during the first morning, and is clearly taking a series of decisions about the construction of her model, which she sees as a vehicle for exploring how her design might be made to work.

Very good performance



"I believe so far my product is going alright, according to plan. The only problem is the lid is not becoming accustomed to the side of the actual product curve of the product, so I need to curve it around to make it less sharp in case of running and exercising. Also I have not yet completed the clip-in for the back. I've only done the elastic band which surprisingly has kind of gone right."

"Overall my product does need a bit of development. Basically I need to complete the building of the clip. I need to develop the lid by adding sides to it to ensure the safety of the pills, so they can't fall out. Also I need to develop the interior to make it comfier for the pills and to keep them from bashing about inside and keep them from rattling all the time."

This student uses her prototype to tackle issues about its shape and form, relating it to the needs of the user in terms of the way in which it might be used. As the design progresses she turns her attention to further problems, such as preventing the pills from falling out, and her awareness of what the product might sound like in use. Her thinking about what the **experience of using the product** might be like is exceptional. She has made substantial progress during the first morning. The audio reportage adds substantially to the understanding of the way in which she is thinking about the design development.

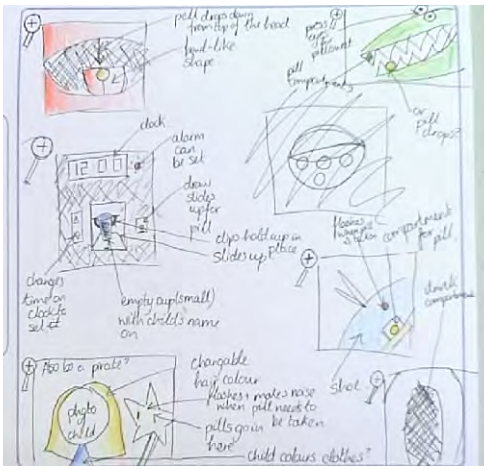
Very good performance

"The things that are working well are the characteristics of the pill dispenser are quite colourful and there are things such as monsters, robots, crocodiles, penguins and fairies that kids would actually want to use. In some there are also drinks so it makes it easier for the child to take their pill".

"The things that need further development are, how the pills are going to be stored and how the pills are going to be distributed to the child. They also need to be more compact maybe so that the child could carry them around more easily."

"The things that are working well are how I've been able to look in detail at how each model is going to work. For example how the pill is going to be dispensed and be caught and how the pill will be taken by the child. I've also been able to see how I can customise the pill dispenser so it can be unique for the child."

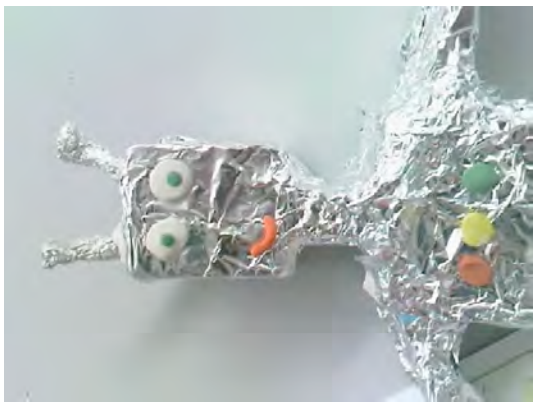
"I really need to think about what a child would really want in a pill dispenser, such as what colours would they like, what shapes would appeal to them. I need to think about how the child would carry round their case, for example a bag or a handle, and I may need to think about an alarm setting."



"I think my designing has gone well so far and my model's quite good, but it wasn't in as much detail as I wanted it to be. I think I've stuck well to the time limits. I like my ideas and think they'd be really suitable for young children, which is who they are aimed at".

"Because I hadn't made a model I wasn't able to see what the buttons and compartments would be like in real life. The time was limited, so I wasn't able to make it as good as would have liked it to be. Also I haven't developed the bag idea yet".

This student spends significantly more time working on box 6, developing design ideas on paper before finally moving on to making a model. This enables her to focus on exploring a much wider range of user needs in her design



Box 15 (Inspiration Pitch)

Students ideas appear to have been inspired from many different sources:

- the handling collection
- the profile cards
- an external source / personal experience
- a 'flash' of inspiration / 'from my head'

These are documented in more detail earlier in this section. The following examples therefore illustrate the quality rather than the breadth of response.

Poor performance

"I'm trying to make it work by putting things together'

'The ideas just popped into my head like always, and I'm hoping to make it work...and I just hope everyone else likes it".

Some students misunderstand the sub-task altogether, or briefly mention a source of inspiration before explaining what they are doing. Other poor-performing students are unable to analyse the source of their ideas, simply saying that their idea 'just came into their head'.

mid - poor performance

"The thing that inspired me was the person's interests."

"I got my inspiration for my idea from a tube in the handling box"

These students have identified a source of inspiration, but have not attempted to explain how it was used as the basis for their idea.

mid performance

"I'm not entirely sure where my idea came from but I based it on one of the cards we were given and I thought it would be quite good to be given exactly the right pills you need when you need them."

"My Nan inspired me because she has a medical box just like this, so I'm determined to make it work properly, and I will do everything to make it work."

This first student identifies one of the profile cards, though does not say which one, but does give a sense of the motivation for his ideas. The second student identifies a relative, and there is a clear sense that he has been inspired to try to design something that will help make her life easier.

Good performance

"What inspired me was actually a pedometer. I realised that if I could put a dispenser inside it, it would work quite well. The idea of the renewable energy inside it - the magnetic momentum - I got that because the battery in the pedometer had run out"

"My ideas came from the props provided to inspire the project. There were many pink and white things which made me think of young girls and there was a jewellery box chair that opened up so I thought I could make mine open up to show something that you couldn't see from the front."

"My ideas came from a James Bond movie. I'd seen this necklace that contained diamonds, so I thought the diamonds could be pills. The other one was from 'prison Break' where they put messages in pens, so I thought the messages could be pills which I thought could work really really well"

These three students make a clear statement of their source of inspiration and go on to explain their thought processes in terms of how the object could be transformed in some way as the basis of their design.

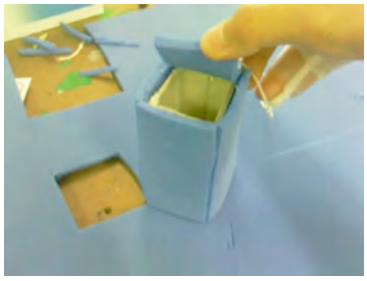
Boxes 17/18 (Day Two Modelling)

Poor performance

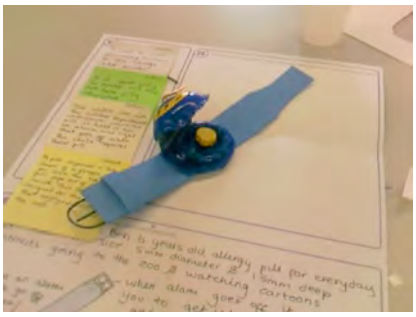
"Nothing's working well because it's not finished. What needs further development? All of it."

"The lid is working well now. I think it all needs development."

This student does make a small amount of progress during the second morning, making the outer casing for his model. However, beyond the fact that it is a container with a simple lid, the audio reveals his lack of thought about any of the other issues involved in the problem.



mid-poor performance



(Box 17) "I think my design is working well as it is nearly finished and it meets the brief we were given, and looks like the design that I've made. I need to improve on my design idea and add to my sketches and notes."

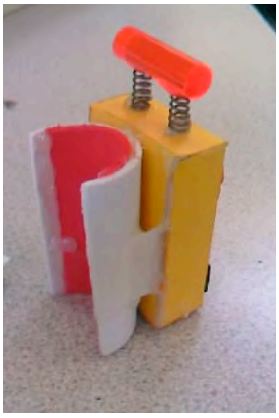
(Box 18) "My product is working well as I have finished my first design. I am now going on to make a second design to show the improvements I can make. I also think my sketches and ideas look appealing and suitable. They also show what work I have done. I am now making an improved design to show the many ways the product can be changed for the better."

This student comes up with a promising solution, and then makes a 'presentation model' of it on the first day. The audio seems to suggest she has ideas for further development, but nothing specific is said, and in practice no further work is undertaken at all during the second day.

mid performance

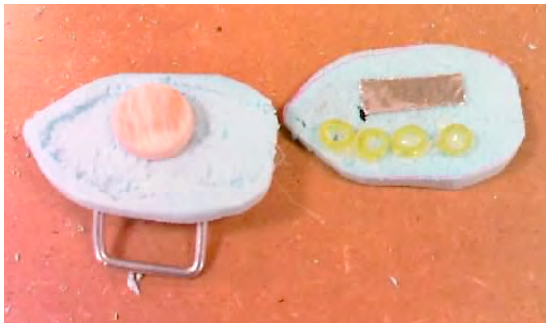
(Box 17) "My product is working well because it's nearly finished. I need to develop the plastics more as I need to sand them."

(Box 18) "Everything is as I said before. I need to show more detail on paper about how it works."



During day two this student has developed a reasonable idea, addressing the need for a range of features and bringing them together into a single container with a dispensing mechanism, and an attachment to a bike. In doing so he has effectively 'black-boxed' the internal physical and electronic workings of the inside, producing something that could potentially be made to work. Unfortunately he makes little progress on day two, putting the finishing touches to his model but avoiding addressing the issues involved in 'how it works'.

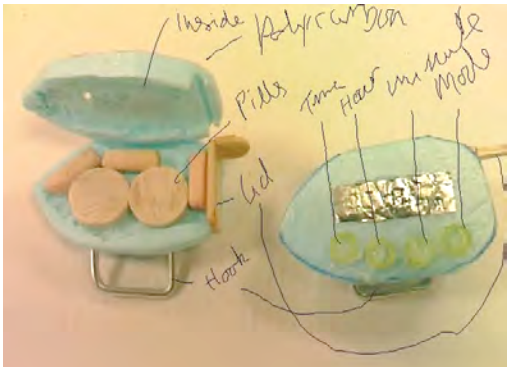
Mid - good performance



"I found out that I could fit quite a lot of pills inside, and that the ones with different shapes could actually fit in. Now I need to finish the actual cut-away so I can see what's going to happen and to see if the mechanism holding them together would work."

"I've found out that I've got quite a lot of space so I can fit the mechanism inside and I've found out what each button could be used for. I still need to develop how I'm going to fix things together and how I'm going to get the pills in."

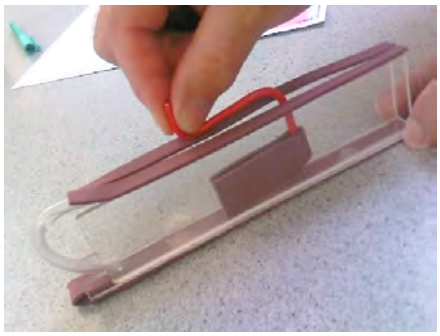
There is a clear sense here that the purpose of the model is to develop the design by being able to test sizes and shapes out in 3D.



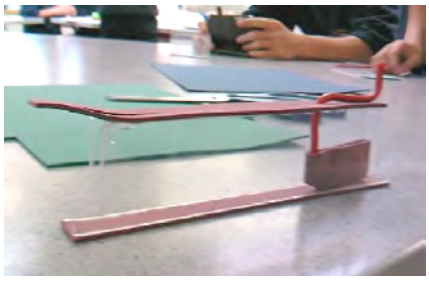
Good performance

"I've got to work on the sliding, and it needs to work a lot smoother".

"My model is working well and I've designed a good slider to push the pill along so that it will come out of one end perfectly."



"My further development will be at one end I will have something to collect the pills as you pick it up at one end."



At the end of the first day the student realised that the model he was making to explore the overall shape and form of his design would not enable him to develop the mechanism he had conceived, so begins a new model at the start of day two. This time he works at a larger scale in more resilient materials, and focuses only on the mechanism, and not on the casing. Unfortunately his audio comments are lacking in explanation of his decision-making processes, and he does not have time to complete the second model.

Very good performance

"I believe my product is going well because the gel inside is padding the interior of the product very well and will act as a cushion. I've tried rattling the pills around and it seems to be working alright. Also I have now produced the clip on the back and also the elastic band so you can apply it to your body. I've also done the pedometer and watch display on the front."

"The actual design needs further development. The shape of it is basic and narrow with sharp edges so I need to curve that off so it's easier so for the purposes of the product so while you are running or exercising you need it not to stick into you. I need to develop that to make it more efficient for the body so it doesn't stick in. Also the basic design is a bit boring so I need to add colour and stuff."

This student continues to develop her design throughout the morning of the second day, using the model to test out her method of reducing the noise made by the rattling pills. Her analysis of what still needs to be tackled shows an unusual awareness of ergonomic issues of how the container will fit the human body, and of the need to improve the appearance. Given that there is very little further development recorded on paper, her audio comments add substantially to the evidence of the quality of her thinking.

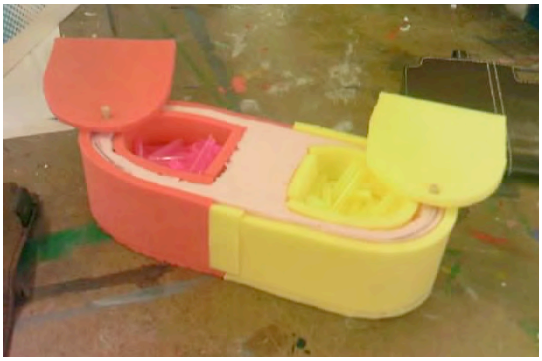
Box 19 (Video presentation)

The lowest-performing videos tend to be disorganised, with the product failing to work and the audio obviously un-prepared and incoherent. The time (30 seconds) often runs out before the presentation has been completed.

Below average performance

"You press this button - the one that is sellotaped on and a pill falls out that I forgot to put in, and oh yeah, there's this thing that makes it look good, and oh yeah, it's green, which makes it look cooler. Oh. Dropped it."

"This is my pill dispenser. It's in the shape of a pill. You open it like this. The red pill goes in there and the yellow pill goes in there. And that's my pill dispenser."



These presentations show models made with widely varying degrees of skill. However, the short commentaries are unhelpful in explaining how the idea is intended to work and who it is aimed at.

mid performance

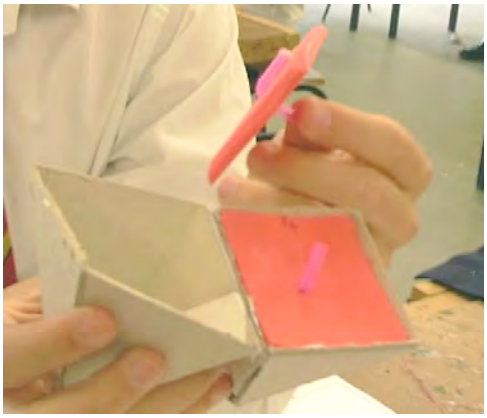
"Basically what I've got here is a pill dispenser for children that you can wear. But it's not any old pill dispenser. What you can do is press the button here and like a tic-tac box, out pops the pill at the bottom. So it's simple to use for children."

"This is the button here you press to start the four hour timing. This tells you what the time is, and this speaker sounds the alarm. Here you twist it to get to the yellow bit and after the alarm goes off you twist it to get the red pill out."

The first dispenser is shown 'in use' and the target market is mentioned with a simple but clear description of how it works. In the second example the dispenser is held in the student's hand and its operation fully demonstrated in the video clip.

Good performance

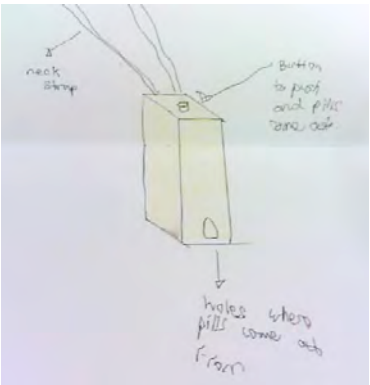
"This is my design, modelled by Lucy. You can see how it works inside, it's coiled around the piece of plastic so you can re-fill it. That's the timer, so that you know when to take your pills. You can rotate the side bit to make the pills come out and just tear them off. Bye bye everybody!"



"This is my tablet dispenser. As you can see I've chosen a shape with two triangles. Here is a latch that keeps the triangles closed together. When you open it up there are two compartments, so you can keep two separate tablets. The reason I've chosen this is that I've designed it specifically for a 16-year old and I thought it would be attractive and interesting."

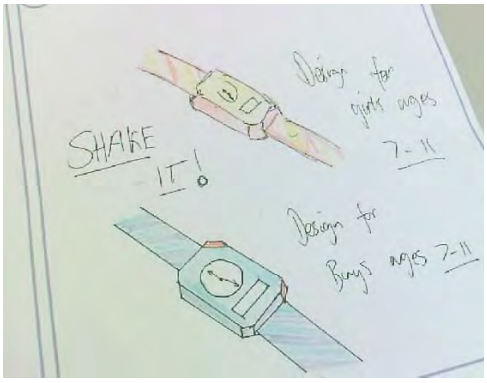
The best videos demonstrate an element of 'sparkiness', and their purpose and audience have clearly been considered. In the first example above the dispenser is held by another student to demonstrate as the lively and light-hearted but concise description of its operation proceeds. In the second the student stands in the corridor outside the studio, deftly demonstrating and describing the operating device and identifying the user all within the 30-second timeframe.

Box 24 Fast Forward



Poor performance

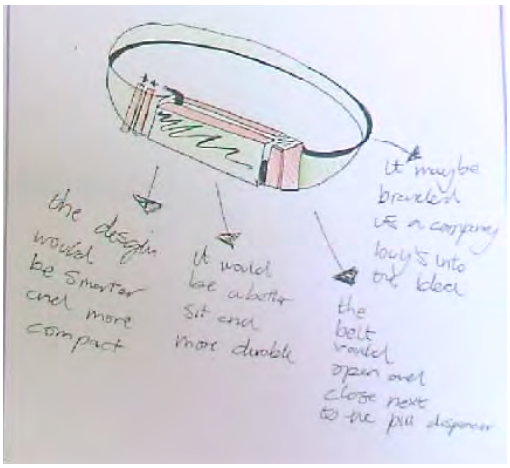
The weakest responses tend to do little more than re-draw what has already been made, re-iterating how it works.

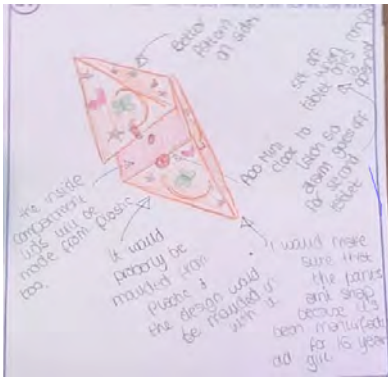


The student on the right has managed to go a bit further and has represented two different watch-based designs simplistically differentiated for boys and girls.

Mid performance

The student on the right uses Fast Forward as an opportunity to start to explore some of the wider issues in his design after having focused almost entirely on technical aspects during the modelling. As such though he is identifying what still needs to be considered rather than proposing possible solutions.

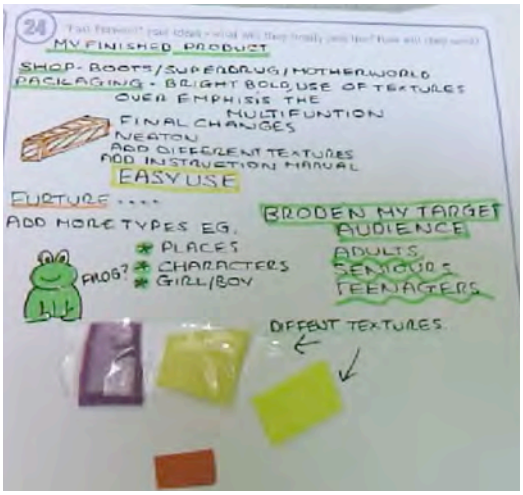




Meanwhile another student uses the opportunity to develop the pattern on the outside of her container, and also starts to consider the materials that might be used in its manufacture.

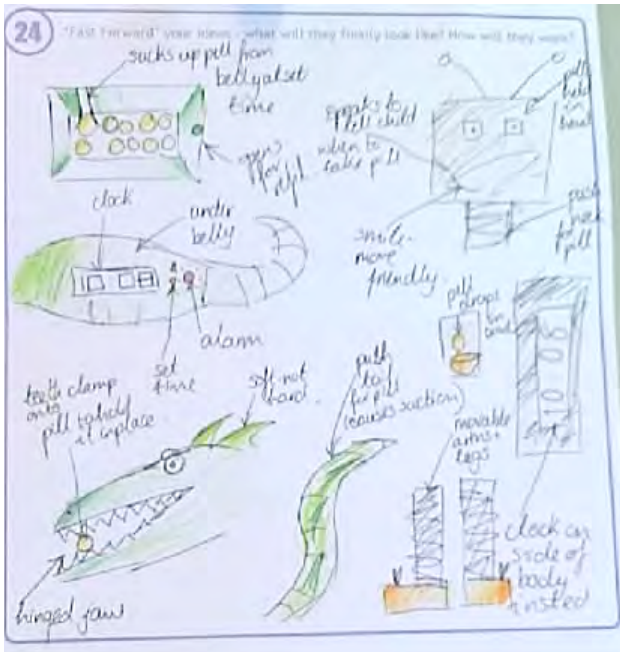
Good performance

This student extends her design thinking in terms of packaging and marketing and adding extra features - including an instruction manual, and finally has the insight to include possible sample textured materials.



Very good performance

Meanwhile this response is clearly full of lively and creative ideas for the further development of new features and how they might work, as well as changes to make the product more user-friendly for a child.



Box 25 (Lift Pitch)

Poor performance

"This is my pill dispenser. It works well and it's an innovation in the world of pill dispensing contraptions".

"Right, I'm going to try and sell my pill dispenser to you. So, like my pill dispenser is top-of-the-range. It's easy to do. It's stylish. Everyone loves it. OK? Thank you listening to me. Hope you buy my pill dispenser."

Although these pitches do include some appropriate phrases, they miss many opportunities to make more of their product's particular features and benefits.

Mid-poor performance

"Basically this product has never been on the market before. It's good, it's stylish, young people would like it, those who are on medication for a long period of time, they could have it then. It's got batteries in. It's all you want, it's hand-held. It's easy and it's a general good product that's never been on the market".

"You should buy this box as it is very secure for little kids' pills so they will not lose them. It has compartments in it to store up to six pills and has funky pictures on it so the kids will like it. It has a hinge on it to stop the pills falling out, and you should buy it because it will make you loads of money"

These pitches are enthusiastic and lively. However they present the features in a somewhat haphazard way, though the latter pitch does address the issue of the potential return on investment, which few others do.

Mid performance

"My pill dispenser is unlike any other pill dispenser. What I have done with it is I have made a perfect design so even those who are old with arthritis and bad a grip can still hold and operate it perfectly. All you need to do is grip it in your left hand, push a button and the pill will pop out straight into your hand. I have also added extra foam and padding in the exact places where you need it to stop irritation"

This pitch focuses well on the way in which the design meets the needs of a particular user, and takes the opportunity to 'demonstrate' the simplicity of use of the product.

Mid-good performance

"So tell me, are you sick and tired of carrying pills around with you everywhere?" "Yes!" "Do you have to carry those stupid foil packets round with you everywhere?" "Yes!" "Well now you don't have to. With the Pendant Pill Dispenser you just put the necklace round your neck and you'll look cool and trendy in front of your friends and still have the stylish look. Then you can take the pills if you need to and just forget about it. It's great!" " Really?!"
"Yes - it's all it takes!"

(Whispered) Let's be quiet about it so no-one else will find out. It's a revolutionary design with a simple but effective design enabling the person to take their pill easily, brilliant for extreme conditions and sports.

These two pitches are particularly imaginative. The first is a conversation between two students and would work very well as an advertisement on the radio, but as such somewhat misses the point of selling to investors - though might have formed an attention-grabbing opener for a more serious business pitch. The second is very short, but the implication that the idea is so good it needs to be whispered is 'marketing genius'!

Very good performance

"It's appealing to all ages as the design and colour can be changed on the outside. There's a tape-like dispenser which makes it easy for people with small hands or who have difficulty seeing. There's an alarm to be used when people don't know when to take their tablets. It's got a twistable end which makes it re-fillable, so can be used again and again, which makes my design have a lot of potential."

This student successfully presents an organised and coherent sequence of clear user features and benefits.

"Hello, may I have just a second of your time to explain this new pill dispenser and container idea? It's a unique product - I've never seen anything like it on the market. It's very clever, such as the lid that lifts up electronically. It tells you when to take the pills due to a display and timer and you can set it to what time you want each pill. It's stylish and would fit any modern contemporary home interior."

"Hey well, I didn't know I was going to catch you here. I'd just like to talk to you about some designs we've been working on for years. We've finally perfected it. We've got all the right harmonious colours, suitable for teenagers. It's safe. It works like a charm you know - it's got an amazing dispensing mechanism. We really think it's going to hit it big. It's cheap, but not too cheap. Practically indestructible."

These pitches both effectively grasp the concept of being in a lift with potential backers. They are lively and enthusiastic about their products and emphasise a number of key features, incorporating key words and adjectival phrases to denote uniqueness and sales potential. Note the considerable sophistication of 'It's cheap, but not too cheap'!

Final Review comments

Poor performance

Many students made either no or very few review comments. This was generally due to a lack of time at the end of the second morning of the task.

Mid performance

We took three pictures of our design and after that we recorded our voice and how we could improve and what was good about it.

We had to do a video and explain what we made and how it worked

We had to say three good things about our design and three things we could improve

This student simply describes what he was asked to do without making any comment.

Good performance

If I had more time I would have further developed my ideas and maybe come up with more interesting ideas and styles.

My first model if I had more time to spend on it would have had a strong idea and the materials I would have used would have be better and more thought through.

Here the student takes the opportunity to consider how things might be done differently if there was more time.

Using the booklet as well as the tablet was a really good experience. My idea first started to develop when I could put it on paper.

This helped me think more about my design and about what my client would really want.

Taking pictures was good so I could see how my model developed along the course. Also making a voice record was much better than writing things and it was better for me because I feel more comfortable talking than I do writing.

Meanwhile this student is starting to analyse the process in terms of the impact the different actions had on the way her ideas developed and were recorded.

Very good performance

Good points made although could use some more information and things that the product will require

Improve on voice recordings, making it clearer and easy to understand

Good photos of product and work so far even though it was not developed very far Could use different angles on camera.

In this review the student identifies aspects of the process that could be improved, including the way in which she took her photos.

Here I didn't focus on one thing, I should have focused on the lid and found a way of attaching it instead of doing other stuff.

I needed to realise that it was too big and I needed to make it smaller before I could go on.

This student has identified the moments in the development where on reflection she would have benefited from doing things differently.

summarising e-scape performance characteristics

Top portfolios

The highest ranked students tend to develop a straight-forward solution that focuses on and incorporates a working mechanism of some sort. They often make more than one 3D developmental model to help them resolve aspects of their idea. Their designs are generally convincing as potential products that would be distinctive in the market place and might actually be made and sold. Other high-ranking portfolios present good annotated graphics and produce solutions that would appeal to the specific market they have identified. These portfolios have all reached some sort of 'conclusion' that is presented in the video, and they have generally worked steadily through the task to get there.

Upper mid portfolios

Moving down the rank it is more difficult to discern clear patterns of response. Different learners may well produce work of quality in one or two respects, such as perceptive user-consideration, or perhaps a lively inspiration, video or lift pitch file, but tend to lose ground due to their product being too complex or large, too 'wacky' to be realistic, or being unresolved at the end of the session.

Lower mid portfolios

After about the 250 rank order position the content of the portfolios tends to become more simplistic, with students often making simple rectangular boxes with compartments and record short and/or less appropriate responses to reflective and analytic steps. They demonstrate limited enthusiasm as they 'plod' through and concentrate on presenting what they see as their final design rather than developing it further. However there are some exceptions in this range in terms of unusual but unresolved ideas.

Bottom portfolios

Portfolios at the bottom end of the rank order have tended to concentrate more on the general appearance and appeal of their product, rather than how it might work. The extent of their development is often extremely limited. They typically conceive and attempt to construct a very basic form.

For comparison, the official GCSE performance descriptors for A C and F are reproduced in Appendix 29.

12. the scalability of e-scape

The geography and science strands of work were devised explicitly to address the ***transferability*** element of the brief for phase 3. Is the system that was originally developed within the context of assessment for design & technology applicable and appropriate for the assessment of other subjects – in this case science and geography? The answer – as section 10 shows – is that it is.

The other element of the phase 3 brief concerns ***scalability***. Is the system capable of being scaled up from a research activity run by a research team to a full national assessment system operated by Awarding Bodies and schools? So, whilst phase 3 remained within the context of a research and development operation (i.e. it was not being run by an Awarding Body), we have sought to gather data that will inform the scalability questions.

It seems appropriate to tackle the issue in three parts:

• the scalability of the task-generation process in the authoring tool
• the scalability of teachers running the tasks in schools
• the scalability of the assessment process using the pairs engine

scalability of the task-generation process in the authoring tool

It is important to note that the authoring tool is a web-based tool that is ubiquitously available provided one has the appropriate URL, username and password.

The use of the authoring tool by teachers and others during phase 3 has been reported above and summarised in the section on Research Questions that follows below. Essentially all users found the system easy to operate and as a result of their combined activities it has approx 100 tasks logged into it … including:

- bungee-jumping
- bow tie for teacher
- a day at the beach
- celebration
- fish lorry
- shadow puppets
- bird scarer

This profusion of tasks speaks to the flexibility of the authoring tool and the success of our teachers in creating tasks through it. The only technical challenge to scalability is a usability one (the extent to which the software and its functions is intuitive). This is a matter that is being kept under close observation and development of subsequent iterations of the software will no doubt become ever more usable.

In phase 3, the training for using it has generally been in the order of half a day, but we have also been developing user guides and support material that will be available on line and that will therefore enable us increasingly to undertake most – if not all - of this training remotely.

the scalability of teachers running the tasks in schools

This is probably the most complex area to comment on since schools are so different and teachers' confidence in dealing with digital systems in the classroom remains low. However there is data to inform the issue from the perspective of our phase 3 teachers and this is reported above in detail and summarised below in the section that follows.

Most teachers received a one day training session in using the system and as a result most were 'confident' or 'very confident' that they could operate the system effectively. There were some who were 'a bit unsure' about it but in the event - when it came to running the test - in response to the question "How easy was it to manage the activity via the laptop" 15 out of 16 responses were that it was 'easy' or 'very easy'. Moreover all 18 responses to the question "How easy was it to cope with the handheld computers" were that it was 'easy' (10) or 'very easy' (8). From the standpoint of user-manageability therefore, it seems reasonable to suggest that the system is scalable.

However, there is a different aspect to scalability here since the system we were operating required class sets of hand-held kit (Ameos and Toshibas) to be delivered to each school in the sample. This is clearly not a scalable system since it could not be replicated nationally, and schools may well choose not to invest in this kit for themselves. It is for this reason that we have been developing the **e-scape-on-a-stick** solution.

Essentially this enables the e-scape system to be operated from a USB stick with each learner having their own stick. When undertaking the activity, learners put their stick into **any** device that the school can provide for them (mac / pc / linux : laptop / desktop / hand-held) and they can use its facilities (e.g. photo / text / drawing / sound / video) to undertake and capture their e-scape work. If the device into which the stick is inserted has an internet connection then the learner work will be seamlessly uploaded into their web-portfolio. If the device is not connected, it will save all work to the stick until it senses that it is in a connected device – at which point all work will be synced into the web-portfolio.

The low cost and high capacity of USB sticks makes this option eminently possible, and the system was developed to a beta version by the start of the summer term 2009. We are currently running a series of trials (and resulting modifications) to the system and it will be available in September 2009 for the launch of the Awarding Body pilot. it will render the e-scape system far more scalable, since schools will not have to purchase special kit, but can use whatever kit they have available. Of course this does not prohibit the acquisition of additional kit, but it does not require it.

the scalability of the assessment process using the pairs engine

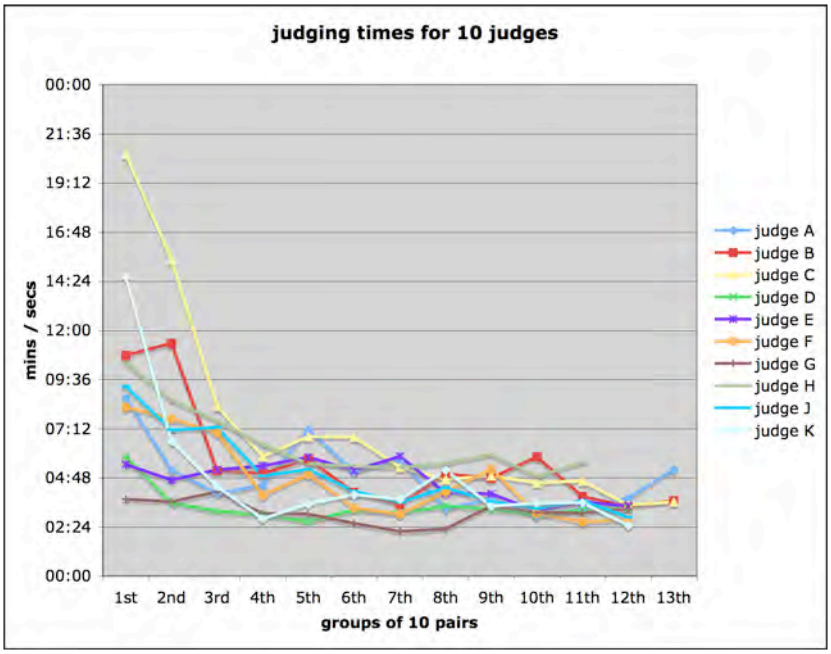
As with the authoring tool, the pairs engine is a web-based tool that is ubiquitously available provided one has the appropriate URL, username and password. To that extent it is readily scalable.

The judges in phase 3 have all reported how easy they found it to use the system, and Pollitt has reported the measurement efficiency of the system (see above). These are summarised below in the section that follows.

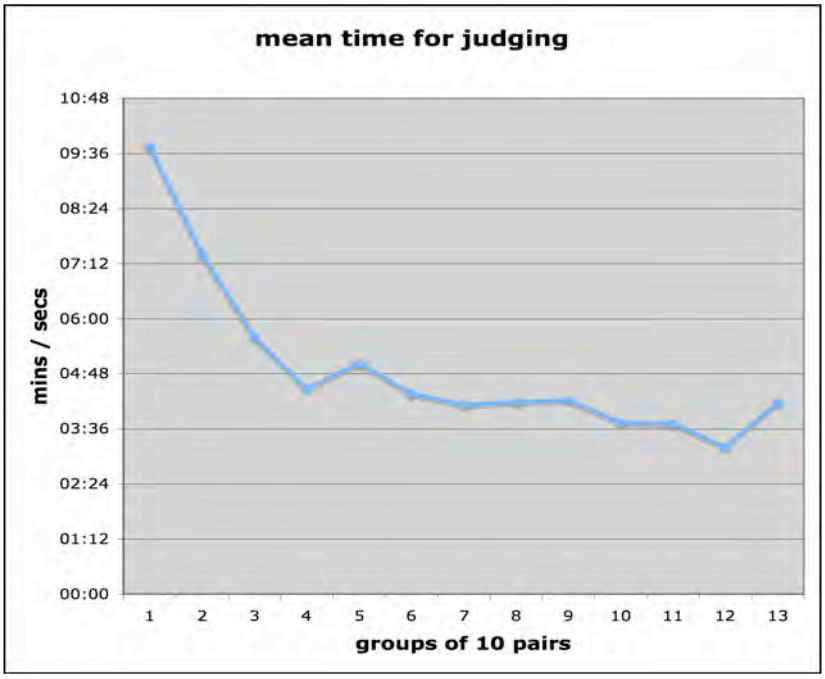
The one day of training that we designed for undertaking pairs judgement can readily be reduced. As with the authoring tool, we have developed a user guide that could be on-line and operated remotely. As part of our trials, we have undertaken a judging exercise with a group of learners who had undertaken the activity. They learned the system and each made 20 paired judgements in half a day. There seem to be no technical difficulties that present any scalability problems with the assessment process.

However, we are repeatedly asked how long it takes to do the judging, and we have collected this data for all our judging team of 28. They report to us that they find the holistic judgements that we require of them are easy to make and professionally more valid than other existing systems – but the question remains whether it is time-effective.

To address this question we have monitored the time taken by judges to complete groups of 10 paired judgements. The first group of ten takes longest as the judge is coming to terms both with the portfolios and with the pairs engine interface. The 2nd group is typically quicker and the third quicker again.



By the 3rd or 4th group, judges have typically reduced their judgement time to between 3 and 6 minutes, and by the 8th group they have reduced it further to between 2 and 5 minutes. The mean for these 10 judges is shown in the chart below, and the median time for making a paired judgement – across the whole of the 130 judgements and across the whole judging team – was 4 minutes 6 seconds.



This means that each judges' allocation of 130 paired judgements took them approx 8.5 hours. Based on our work in phase 2, we had estimated 10 hours, but the quicker process in phase 3 is undoubtedly explained by the efficacy of the new pairs engine.

The scalability question is how this compares to the kinds of GCSE coursework assessment that teachers are doing currently.

We had 28 judges and 350 portfolios, and one might describe this as 13 portfolios per judge. In the current coursework arrangements teachers are responsible for assessing the work of their GCSE group and if we assume that this is 20 learners, then the 8.5 hours that our judges took to do 13 portfolios, rises to 13.3 hours for the group of 20. However, our judging operated through 20 rounds of judgements (i.e. each portfolio was compared to 20 others) and this is far in excess of the necessary level of judging and produces a level of reliability that is much higher than current systems can produce. Pollitt's current judgement is that after 12 rounds the reliability remains at least as high as current arrangements. If 20 comparisons per portfolio for a class of 20 takes 13.3 hours, 12 comparisons per portfolio for a class of 20 should take 8 hours, which is approximately 24 minutes per candidate. The question is whether it currently takes teachers more or less than this time to assess the portfolios of their GCSE candidates. In the words of one AST, a head of department in our judging team:

> With conventional GCSE portfolios it has, in the past, been quite a "painful" experience doing the marking. Usually the first few can take up to an hour each for the larger (better??) ones and reducing down to about 20 or 25 minutes as I "tune in" to the marking criteria. I have also spent quite a time (an hour or two) pre-reading a few folios to get a feel for their overall standard and a rough rank order. Of course added to this can be a few hours (3-4) of internal cross moderation when there is more than one specialist option or more than one teacher marking work from the same exam board.
> A group of 20 folios including internal moderation and administration can therefore easily take 15+ hours... Ok, I am quite methodical, but I do have quite a lot of experience as well!
>
> As to which I prefer.... No contest! E-scape judgements win hands down. The time taken is dramatically reduced for the marking; there is no further administration to do or internal/external moderation. I would also have the added benefit of seeing what has been produced by other schools, something normally only available to examiners and moderators... a great bit of CPD!
> (DW by email 11th Mar 2009)

Interestingly, while he expresses the opinion that the time taken for e-scape judging is "dramatically reduced", the reality is that script marking times are much the same – at about 24 minutes. But it is all the other administration and internal cross moderation activity that increases his estimated time to 15 hours.

Scaling up the e-scape assessment system could therefore be envisaged as follows. A teacher entering a group of candidates for e-scape assessment has to agree to doing a specific number of paired judgements. The portfolios from all schools/centres (nationally) are placed into a single national pot and managed by the pairs engine which distributes pairs to the teachers. As DW points out above, one significant CPD benefit here is that ALL teachers see a spread of work that represents *national* performance, not just the performance of the students in their school. Once the judgements are completed, the Awarding Body then manages the awarding process internally.

the problem of missing data

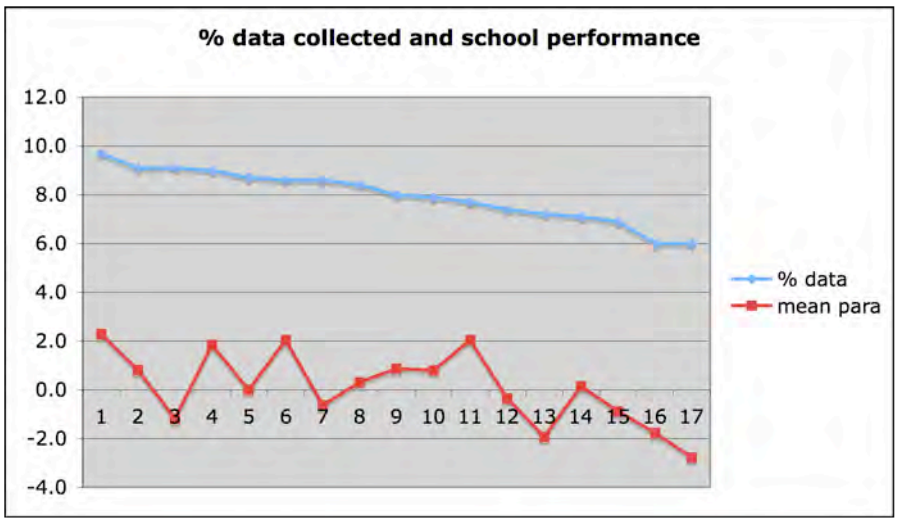
An additional aspect of the scalability question that we have been concerned to monitor is the extent of 'completeness' of the portfolios and - conversely – the extent of the missing data in them. There are two aspects to this problem.

First – we would expect to see some correlation between the completeness of the portfolios and the performance that they exemplify. This is the same in any examination. Scripts with a lot of missing answers will typically do less well than scripts with all the answers in place. Second - and particularly in relation to the scalability issue - we are interested in **why** the data is missing. Has a decision been made by the teacher to miss out a complete section of the activity? Has an individual learner just chosen not to provide any data on a particular box? Has the system failed technically – resulting in not recording a learners' work for a particular box?

We have begun to interrogate the data to find answers here, and it is clear that there is a relationship between the completeness of the portfolios and the mean performances of schools. 'Completeness' here refers to the percentage of boxes in the portfolio with recorded data. A portfolio with recorded data in all boxes would be 100% complete.

The chart below superimposes two scales on the Y axis. The blue line shows the % completeness of the portfolios from each school. So school 1 has portfolios that are 97% complete while school 17 is only 60% complete. The red line is the mean performance of the same schools expressed as their mean parameter value in Pollitt's final rank. It is clear that there is a relationship between the lines – if not an absolute one. School 1 has the most complete portfolios and the highest mean parameter value. School 17 has the joint lowest completeness and the lowest school mean parameter value.

The overall correlation of the two ranks is 0.66, which is sufficiently strong to establish the connection between the two features.



The ***source*** of incompleteness is a critical next question. Much of the incompleteness appears to have arisen from teacher decisions to re-order the activity – e.g. in relation to the timing of the school morning. It is commonly the case that box 9 and 10 show 0% responses from a whole school. This is not just an occasional learner choosing not to do something – it's a strategic decision imposed by the teacher on the whole group. Box 9 and 10 are the photo and sound file recording sessions in the midst of the first morning modelling time. Some schools appear to have decided that two of the photo and sound file sessions fits better into the morning (including break time) than the three sessions that we had designed.

This is by no means a critical matter since – whilst it is true that there is data missing – it is also true that the students will have had more modelling time **between** the two sessions than would be the case in schools where they have three of these sessions. It means only that the dip-stick monitoring process is slightly less tight in the 2 session schools than it is in the 3 session schools.
This kind of decision-making by teachers appears to account for a large percentage of the missing data. In the worst-case school, if we ignore the missing data where the **whole group data** is missing (i.e. where there is 0% data), the portfolio completeness percentage rises from 60.57% to 89%. In effect this is saying that virtually 30% of the missing data **is attributable to teachers decisions** about re-structuring the activity, and 10% is attributable to individual learner variability.

We are aware however that there are some cases where whole school data shows 0% for a box – when our observation notes report that it was done, and where it did not involve box 9 and 10. This points to technical failure in the system – and we are in the process of working up a formal procedure (and associated software checking system) that will allow teachers to scan the completeness of the data capture dynamically as it is happening through the activity. This will be a critical innovation for the developments in phase 4 of e-scape.

13. answering the research questions

The three research questions concern the creation of Structured Coursework Assessment Activities (SCAAs), the operation of SCAAs in the classroom, and the assessment procedures relating to them once learners have created their portfolios. We shall take these in order and review them in relation to the data that we have derived from phase 3.

RQ1. Concerning the creation of structured coursework assessment activities

RQ1 (i) • can teachers make effective use of the activity authoring tool to create their own SCAAs?

The research design involved only a small number of teachers creating their own activities. These were the six 'hot' school teachers that were trained in December 08.

Subsequently we also trained the geography and science development teams to build their own SCAAs using the authoring tool.

There is no doubt that the 'hot' school teachers (5 secondary and 1 primary) had no difficulty in getting to grips with the authoring tool. They each created activities that they felt could be accommodated into their own curricula – incorporating digital resources as necessary.

The authoring tool (in the e-scape website) currently has approx 100 tasks logged to it … including:

- bungee-jumping
- bow tie for teacher
- a day at the beach
- celebration
- fish lorry
- shadow puppets
- bird scarer

This profusion of tasks speaks to the flexibility of the authoring tool and the success of our teachers in creating tasks through it.

RQ1 (ii) • can SCAAs be created – and worked effectively – in science and geography?

Again the answer to this is unquestionably 'yes'. The development teams each created several different tasks and then modified them extensively as a result of their trialling experiences in schools. And again the evidence is in the authoring tool site, with tasks such as…

- re-branding Porthcawl
- eco-day
- Cheddar traffic
- saving the rain forest
- road safety
- reaction timing

RQ2: Concerning the operation of SCAAs and the construction of web-portfolios:
RQ2 (i) • can SCAAs (with all the associated technologies) be operated autonomously by teachers in their schools without the support of the e-scape research team?

There are several aspects to this question – and first it is important to recognise that this question applies to ALL teachers in phase 3 of e-scape: 17 in d&t and more in geography and science. The data to inform our answer is in the teachers' questionnaire. This establishes that (initially) our teachers typically had 'limited' or 'no' experience of working with handheld technology in the classroom or of controlling it through the use of a local area network.

Two requirements of the SCAA were for the teachers to create handling collections and modelling kits - and neither is common practice in schools. Following the training session, 17 of the 18 responses were that they were 'confident' or 'very confident' about creating the handling collections and again 17 out of 18 were 'confident' or 'very confident' about creating the modelling kits.

Concerning the digital kit involved in the SCAA, after the training session, 11 of 18 responses were 'confident' or 'very confident' that they could handle the laptop/handheld system in the classroom with the remainder feeling 'a bit anxious'. But in the event, in response to the question "How easy was it to manage the activity via the laptop" 15 out of 16 responses were that it was 'easy' or 'very easy'. Moreover all 18 responses to the question "How easy was it to cope with the handheld computers" were that it was 'easy'(10) or 'very easy'(8).

At the end of the questionnaire we invited teachers to reflect on whether they would like to use this digital approach to GCSE assessment activities. "Given the chance to use this digital approach within GCSE would you choose to?" And interestingly 11 of the 17 responses said 'definitely', with the remainder (6) saying 'probably'. Among the most common reasons offered for this were:

- students enjoy the interface
- keeps students on task
- easy to record / mark
- innovative
- access to students who are not as academic
- creates an additional dimension to their work
- to record development of ideas - coursework

We also asked if they would like to use the system as part of their normal work in lesson-time. "Given the chance to use this digital approach within regular D&T lessons would you choose to?" Interestingly, 12 of 16 responses said that they 'definitely' would, with 2 'probably' and 2 more 'not sure'. The most common reasons for this were listed as:

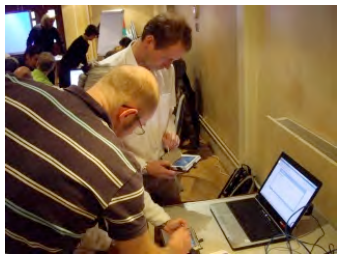
- to focus student learning
- to make assessment accurate / easier
- to enhance collaboration between students

- good variety of information being collected about student work
- in design & development - to record progress / ideas
- onus on pupils
- student motivation
- enjoyable experience
- better responses from students
- wider medium for students to express their ideas through
- instant recording of work
- captures thoughts & ideas which might otherwise be missed

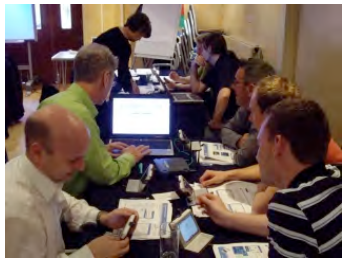
Taken together, these data seem to us to be unequivocal evidence that our teachers could and did run the activities (and the digital kit) effectively in the classroom without our support.

RQ2 (ii) • what degree of training is required by teachers to accomplish this?

In the case of the cold schools – which were the majority of the teacher group involved in phase 3 – they received two kinds of training. Initially we visited the regional groups and spent half a day (typically an afternoon) working with them to outline the requirements of setting up the classroom / the handling collection / modelling kit and the script for running the activity.



Then, much closer to the activity, all teachers came to London for one day of technical training. This involved teaching them to set up the hand-held devices, linking them to the laptop and router wi-fi connections, and working their way through the steps of the activity.



RQ2 (iii) • how does learner and teacher familiarity with the technologies influence the operation?

As we have pointed out above, none of the teachers was familiar with the use handheld technologies in the classroom – or of controlling them in the e-scape manner. They learned the system from scratch and coped very effectively.

With learners the issues were somewhat different. They are familiar with the world of mobile technologies and comfortable with using them. However, few of the learner group had ‘smart’ phones with internet connectivity, and none had experienced the specific machines that we were using (the Ameo and the Toshiba) but they adapted to them very quickly.

As we have reported above, of the 212 students who completed the Ameo questionnaire, the vast majority were very supportive of its value as a designing tool.

- 97% agreed or strongly agreed that the device was easy to learn to use
- 87% that it was easy to use to develop ideas
- 97% that it was good for photographing emerging work
- 92% that it was overall a good tool for designing
- 95% that it was fun to use

And of the 41 students who completed the Toshiba questionnaire…

- 98% agreed or strongly agreed that the device was easy to learn to use
- 98% that it was good for photographing emerging work
- 95% that it was a good text tool for explaining ideas
- 95% that it was fun to use

We have discussed above (section 4) the differences between the two pieces of kit in terms of their value in this classroom setting.

(iii) • how do the two principal technologies (Red-Halo / MAPS) operate in parallel within the classroom/school … and remotely using 3G communication?

One of the main technical issues regarding any system running between a server and client is the management of data transfer from one to the other. This is especially the case for e-scape phase 3 when the clients were Windows Mobile devices and the server was the MAPS / EMS system. The system had to be able to ensure that data created on the device was reliably transferred to the server and, in the case of learner-collaboration sessions, be able to distribute files from the server back to the device.

The solution involved the introduction of Handheld Learning's Red-Halo application. Red-Halo has a small client application that sits and runs on the Windows Mobile device. In most cases this works silently in the background but was available to be scrutinized if problems occurred. The Red-Halo client was pre-programmed with the location of the Red-Halo server. The client was able to perform two functions; first to act as a message layer between it and the Red-Test application, and second to manage file synchronisation.

File synchronisation was achieved by the Red-Halo client examining the contents of the Windows Mobile device and sending the Red-Halo server a list of files and their appropriate details (time of last update, size, etc.). The Red-Halo server was then able to examine its own version of the files on the server and determine which (if any) required to be updated. A series of file upload commands were then sent back to the client.

The Red-Halo client worked in parallel with the Red-Test software, also running on the Windows Mobile device. Red-Test was the interface the student used to carry out the activity. Every 10 seconds the Red-Halo client would poll the Red-Halo server to request a message update. Any instructions to be carried out by the device would then

be downloaded and processed by the Red-Test software. These message were of the nature “start the session”, “move to box 4”, “pause session” and so on. These control messages were generated by the MAPS/EMS system and then sent to the Red-Halo server for storage.

The only difference between running the system in the classroom and running it over the internet via 3G was the known location of the Red-Halo server by the Red-Halo client. When operating the in classroom the Red-Halo client was set to look for the Red-Halo server on the local wireless network. When running the system over the internet, the Red-Halo client was set to look for the Red-Halo server through a fixed URL (web address). Although these operate as separate systems the machine configuration and the software running on the servers were identical.

Although both systems were tested and deployed within a development context, the practicalities of attempting to run a class set of 21 devices within a school using 3G proved insurmountable. The 3G signal operates within the current mobile phone network. Every time a device connects to the network it does so by connecting to the closest cell or aerial it can find. Each cell can cope with a finite amount of traffic. One of the main assumptions of the mobile network is that devices are in motion. If devices are moving they are regularly stepping from one aerial to another and thus helping to balance the load across the network. For e-scape phase 3 it became apparent, and was later confirmed by technicians at T-mobile, that one mobile cell was not capable of handling 21 simultaneous 3G connections.

Interestingly this limitation in the national infrastructure was not appreciated by T mobile at the outset of phase 3. A series of trials established that 12 of our devices was the current limit of capacity for a single cell – assuming they are all static in a single classroom. This forced T mobile to re-write their European advertising and sales literature. See Appendix 30 for a detailed description of the application protocols that allow data to be moved between systems.

RQ3. Concerning the *assessment* procedures using comparative pairs judging:
RQ3 (i) • can the differentiated pairs methodology work through a single, self-sustaining, auto-updating programme?

There is no doubt at all that the pairs engine works very effectively. We have described above its operation both in the main d&t sample and with the science and geography teams. In every case the judges had no difficulty in operating the system and it rendered results that are very reliable.

We have also reported above – in detail - the feedback that we received from the judges concerning how they felt about operating the pairs judgement process and the pairs engine interface itself. They were almost entirely positive about it – and two overview comments summarise the general responses of the team of 28 judges.

This is a visionary use of modern technology to support what good teachers of design and technology have always believed in but have had to stretch the examination systems to show. With this, good Design and Technology can be rewarded as a result of (and not in spite of) the examinations (TL)

Over time, this assessment technique would be likely to enhance the quality of practice of D&T teachers. Where current assessment techniques encourage a reductionist approach exemplified by algorithmic approaches to designing and a lot of pretty-but-shallow work, this technique should encourage learner collaboration; stronger independent working; more reflective ability and self-evaluation; and the ability of students to discriminate good design work from poor work. For these reasons in particular I consider that this approach has the potential to be a much improved method of assessing large numbers of students when compared with existing methods. (DP)

RQ3 (ii) • what training is required to enable teachers to operate as judges?

We recruited a judging team of 28 and drew them mostly from England – but with two from Ireland and one from Israel. We have described above the one-day training session that was held in London. Essentially it involved half a day looking at criteria and matching them to portfolio evidence, then the other half day working the engine interface and becoming familiar with the process of identifying winners in each pair.

A month later, at the end of the judging process (having completed their 127 paired judgements) we asked all judges to complete a questionnaire about it that included a section on the training. The following comments are representative of their responses to a question about whether the training had been sufficient and effective.

Yes (RM)
yes – plenty sufficient and effective; instructions documents very helpful too (RW)
Yes! The transition from criteria based marking to holistic judgements wasn't too much of a challenge for me. I think that much of this can be put down to the initial training in making the pairs judgements (DW)
The training day in London was thorough and gave me confidence to tackle the judging. I felt that all aspects from the technical to the criteria for assessment had been clearly dealt with. (AM)
Yes. Also the system was very user-friendly (PH)
Yes certainly (TL)
*Yes, but it still took me some time to overcome ingrained, detailed examination of the folios and being prepared to adopt the holistic judging system. This became much easier a) after having completed a number of comparisons, and b) once the chain started. (DH)**

At the end of this process, Pollit compiled a detailed analysis of the performance of the judging team. We have included his report in full above, and summarising the output, he comments as follows

> Because every single judgement made can be compared to the outcome predicted (with the benefit of hindsight) from the final rank ordering, very detailed monitoring is possible of the consistency of the judgements made by each judge, and of each portfolio.
>
> The most important statistic indicating misfit is the *weighted mean square* (WmnSq). Theory predicts that this statistic should average 1.00, and in these data it does exactly that.
>
> The *unweighted mean square* (UWMnSq) is a similar statistic, but is less stable and particularly sensitive to occasional extreme 'misjudgements'. Three judges show high values on this statistic.

> Overall the amount of misfit seems quite acceptable.
>
> The final scale spread the portfolios out with a standard deviation of almost 3 units. The average measurement uncertainty for a portfolio was about 0.67 units, and the ratio of these two figures was 4.45. This means that the standard unit of the scale was almost 4.5 times as large as the uncertainty of measurement. This means the portfolios were measured with an uncertainty that is very small compared to the scale as a whole; this ratio is then converted into the traditional reliability statistic – a version of Cronbach's alpha or the KR 20 coefficient.
>
> The value obtained was 0.95, which is very high in GCSE terms.

RQ3 (iii) • can the output from the pairs judging process be incorporated into Awarding Body back-end systems for awarding processes?

Within the phase 3 work we have not formally attempted to integrate the pairs judgement data into Awarding Body back-end systems. We have however had meetings with the Awarding Body technology staff and the awarding teams who run those systems and we have been assured that the data could readily be converted from the e-scape scale to the UMS scale that they use for this purpose. Pollitt speculates on the process as follows:

> Like the raw score scale arising from marking exam papers, the scale on which the portfolio parameters are estimated has little substantive meaning. A suitable linear transformation is needed to support its interpretation. Only a linear transformation will preserve the equal interval nature of the scale the Thurstone method creates but, in principle, any linear transformation will do if it is useful, and the principle is the same however it is done.
> For example, suppose a scale of GCSE grades is needed. Two points need to be fixed, and in this case they would be two significant grade boundaries. They might be fixed by looking for portfolios whose quality of performance is deemed to be on the borderline between adjacent grades, but this is not advisable when the new component is different in nature from the existing ones. A better approach is to determine, for a defined group of students, how many are expected to pass each of two grade boundaries, probably by counting how many do so on other components.
> If, for example, it is decided that 20% of the sample analysed here should get at least an A and 60% at least a C, then these percentages are converted to numbers:
> 20% of 352 =70; 60% of 352 = 211
> Lines are drawn in the ordered list of parameters below the 70th and 211th students, and the minimum parameter values that get the higher grade are noted:
>
> | | | |
> |---|---|---|
> | A/B | Ranks 70/71 | 2.16 |
> | C/D | Ranks 211/212 | -0.61 |
>
> The distance between these two points corresponds to 2 grades, so the width of a single grade is calculated as (2.16 – -0.61)/2, which equals 1.385. We can now define all the grade boundaries by adding/subtracting this number as appropriate:

Pollitt applied this process to the 352 e-scape portfolios and the results are shown in his report (see section 10). The exact positions can, of course, be adjusted a little through the awarding process by qualitative checking of the portfolios. Pollitt concludes:

> It is worth noting that the average standard error of measurement for a portfolio was 0.668, which is just less than half the width of one of these "GCSE" grades. It is unlikely that many GCSE components – or any that are judged rather than scored quite objectively – could match this level of measurement accuracy.
> This illustration simulates GCSE grades, and facilitates aggregating the results with other GCSE components, perhaps by means of a kind of UMS. But any number of grades with any desirable properties could be chosen instead, particularly if the purpose is formative rather than summative.

Quite apart from this phase 3 work, TAG Learning have for the last few years been closely involved with a number of Awarding Bodies developing and maintaining

systems designed with the express purpose of submitting exam coursework electronically. In the process they have tackled the technicalities of linking data into AB back-end systems.

It is necessary to enable data to move from one system to the other in a format that can be validated and authorized. This requirement in relation to internet-enabled applications is usually achieved through the publication of an API (Application Programming Interface). In simple terms an API is an agreed set of functions and methods that can be used to achieve the desired outcome. The pairs system already has an API built in to it. There would need to be a technical discussion in relation to any particular Awarding Body to determine whether it would be most effective for them to adopt this API or to incorporate their API into the pairs system.

Another approach, which TAG Learning is currently using with AQA, is a more straightforward transfer of data in and out of the relevant systems through csv files. Every day the submitted and moderated marks are summarized in a spreadsheet file and sent across to AQA using a secure FTP connection. Data regarding centres and candidates is also passed back from the AQA system using the same method.

In either case it will be possible to pass back from the pairs engine any detail relating a learner's work be that a simple parameter value, a rank placing or an estimated projected score

See Appendix 30 for a detailed description of the application protocols that allow data to be moved between systems.

A final reflection on the research questions

At the end of the research questions, we make the observation that …

The answers to the above questions will enable us to answer the key question of whether e-scape methodology provides a sustainable national system of assessment at GCSE.

In relation to this over-arching question, it is worth pointing out that we are now in detailed negotiations with two Awarding Bodies. Plans are advanced for running parts of existing GCSE and Diploma qualifications in e-scape format, starting in September 2009. With each Awarding Body we have identified a qualification and the sections/modules of it that lend themselves to e-scape methodology. We have identified an approximate cohort of schools/centres in each case and are now planning what will be an Awarding Body pilot programme monitored by Ofqual.

This will become **e-scape phase 4** through which learners in schools and colleges will receive live awards on the basis of their work undertaken and captured into web-portfolios using the e-scape approaches developed over the last four years of research and development.

14. an account of the whole e-scape experience in one school

Dave White: Subject Coordinator for D&T, Clevedon School.

The starting point

At Clevedon school we joined the e-scape project at the start of phase three, we were somewhat fortunate in that we were experienced in the relatively new OCR Product Design GCSE. The innovation challenge module being which based upon the early paper based version of the e-scape project gave us a good grounding for the move into the use of Hand-held Technology.

Our background and students

All of our students who were involved in the e-scape project had recently been through the innovation challenge but were still very excited about this new development. The opportunity to take part in e-scape was offered to 30 students from our two Product Design groups; this enabled us to have a group of students whose background in Product Design had leanings towards designing and making products using graphical, resistant or compliant materials. We also invited two year 9 girls who had heard about the project and were so keen to take part that they could not be turned away. As it happened, reducing the 30 plus students down to a group of 21 pretty much took care of itself, as some students were unable to join in due to commitments from other curriculum areas on the day.

Pre preparation

Preparation for the project needed to be well thought through and properly planned. The Goldsmiths and TAG Learning team provided us with excellent guidance on the ethos of the e-scape project, organisation and preparation for the activity, and training in how to use the Hand-held Technology. Throughout this stage I was also able to draw upon my experience with the OCR exam to enable me to prepare thoroughly for the activity. It was clear that time spent in preparation would have huge benefits for the smooth running of the task. Notifying parents of our intent and gaining their permission, organising for the students to be released from lessons and developing resources followed the same pattern as the OCR exam so there were no real surprises along the way.

The handling collection and other resources

Having run similar activities before confirmed in my mind that if students were to design creative and innovative products the "inspiration materials" provided for them had to be very well considered. This handling collection to a great extent governs the starting point for many of the student's projects and it is therefore vitally important to make sure that the handling collection contains a wide variety of materials that spark ideas but do not give the answers. Richard Kimbell gave us a considerable amount of advice and help in putting together the handling collection. Working with other teachers from local schools gave us the opportunity to think about what sort of things would be appropriate. We were also fortunate in that we were able to share between us the contents of each other's handling collections. This collaboration allowed us a greater variety of materials and the ability to share ideas about what would be good to include or exclude.
The handling collections are a great source of inspiration but can be somewhat limited due to the practical considerations of size and expense. To overcome this I put together a rolling slide show that would be displayed using a digital projector throughout the project. It took some time and lots of head scratching to select the images for the slide show, there was little point in duplicating the materials from the handling collection so I concentrated on images of potential clients, mechanical systems, possible scenarios, and iconic products. To provide quality images I made extensive use of the www.flickr.com website and of course Google images. (Tip. A looping .wmv .mp4 or .avi file allows you to run the slideshow in a window instead of full screen as you would have to with a Powerpoint allowing the rest of the screen to be used for other resources). I did consider trying to find some suitable background music but decided against this when I thought about the amount of noise in a small design room... and trying to find music for a project about pills gave some interesting results when searching on the internet! Another really useful on-screen resource was to display a countdown timer, although the tasks in the activity are controlled by the software and hand held devices the students find it really useful to be able to glance at a "clock". I have found that the freeware "Cool Timer 3.6" from Harmony Hollow Software works very well.
http://www.harmonyhollow.net/cool_timer.shtml

Resources- materials

Careful consideration was also needed of materials to be made available to the students for modelling their designs. To be able to work quickly the students need materials that can be worked easily using basic hand tools and simple fixing methods. However the quality of the eventual outcome produced by the student is very much determined by the

quality of materials that they are provided with. Getting the balance right has a significant effect on the range of different ideas that the students are able to model as well as how successful the model will be as a working prototype and also to demonstrate its aesthetic qualities. In selecting the materials I tried to ensure that there were "free form" materials such as modelling clay for quick experimenting, sheet materials such as card, plastazote, or foamex for constructing shell structures and a variety of sticks, dowels and strips of softwood for frames, linkages etc. I also provided a wide variety of carefully selected scrap materials (purchased from our local "Children's Scraps Store"). The selection of the scrap materials is very important; the students find it very difficult to differentiate between junk modelling and the use of scrap materials! The most useful scrap materials are the items such as cotton reels, plastic containers, buttons, threads, fabrics, etc. It was apparent that the students would want to apply surface decoration to their designs, to do this within the time constraints I realized that it would be a disaster to allow them to use paint (in my experience pupils and paint don't mix!!) so I provided offcuts of self adhesive vinyl and a few marker pens. Making sure that all these materials were organised was again a good idea, I tended to group modelling materials into boxes by type (sheet, strip, rod, fabric, etc) rather than by material... it makes choosing much easier and students can select what they want for their product by comparing similar shaped materials side by side and trying out their different working properties.

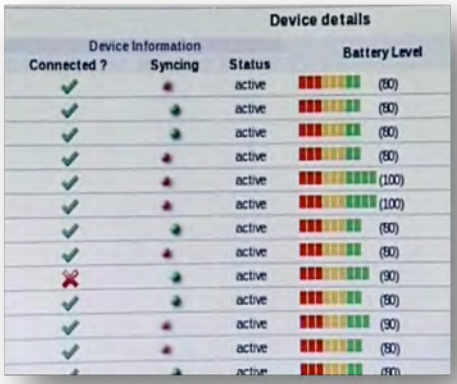
Running an e-scape project... and the script

Smooth running of the e-scape project was critical, having read through the teacher script I realized that the pace of the task needed to be fairly brisk to avoid unfinished work and incomplete portfolios. It is essential not to go into this sort of activity "blind", reading through the script before hand and having time to think about what each stage might entail can save moments of panic when running the activity. Ideally for a project such as e-scape two members of staff should run the activity, one to deal with the technology, and one to direct the students. Unfortunately I found myself alone but having prepared myself for the activity and knowing what to expect I found that it was possible to cope most of the time, even when I had the added complication of the film crew recording all of the events for Becta. The script prepared for the e-scape project has been carefully considered and guides the activity from start to finish. The timings for each of the tasks had been planned to allow students to both engage in the activity and also to record their work for the e-portfolio. Reading through the script beforehand allowed me to not only read word for word the instructions but also to expand on this and explain what was required clearly and concisely. It is a delicate balance to not just read religiously word for word from the script whilst also trying not to influence the students design work. The script should be considered as the "director", but it is up to you to make the task exciting, fun and productive. You need to take control and be on hand to take on roles of "editor", "producer", ...and gofer.

More forward thinking...I was a little concerned that some of the less "forward" students in the group would have some discomfort doing their voice recordings in public... as it turned out most of them got over their reticence very quickly, but just in case I provided them with a booth made out of a couple of sheets of MDF temporarily screwed together and obviously called "the big brother diary room". I also knew from the innovation challenges that we had previously done that when the activity really gets going there is often little space left on the tables to take clear and uncluttered photos or videos. I therefore provided a separate clear space with a plain fabric backdrop and a desk lamp where students could take their product to be photographed. At first few students bothered to use the photo booth but as desks became covered in a deeper and deeper pile of materials more and more of them started to see the benefits of this space.

Hardware and software

As for using the e-scape hardware and software. pretty easy really. The students had a brief training session the day before running the activity.... a great opportunity to let them get over the excitement of using new kit and also to start to get to grips with the capabilities of the text recognition, camera and video recorder. And as everything looks and feels like using a normal windows computer students had few difficulties with the interface. As for running the session from a laptop, this was very straightforward... the only problems came in having to occasionally wait a few moments for the hand held devices to sync and upload the data to the e-portfolios. ... the " jiggling balls" did however show me when it was ready for me to move on to the next task. The practice run also allowed me to the opportunity to make adjustments to the positioning of the wireless router... really important to get this right as all the hand held devices communicate via the router and you need to avoid having areas of the workshop with a weak signal.

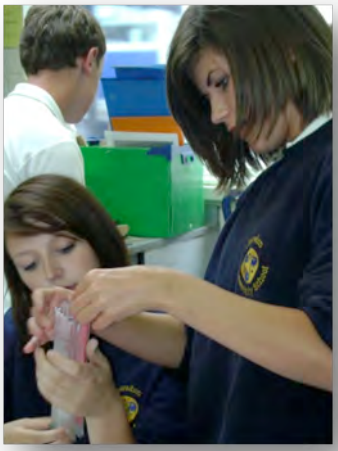


Design stage

The designing stages of the project were again pretty straightforward... anyone who encourages students to use design skills from the KS3 strategy (4x4, scamper, etc) should have few difficulties here. Our students were pretty familiar with sharing design ideas in a collaborative way... e-scape just made it easier! My only regret is not having assigned the students their own individual colour to use when suggesting modifications or annotating their team's work. It became difficult to differentiate who had contributed to others and who had taken note of other student's suggestions.

The students liked being able to demonstrate their design thoughts using a variety of methods. Whilst some preferred sketching others found that they could communicate more effectively by making a quick model and then photographing it. Many liked the way that the task was broken into manageable chunks of time with opportunities to review their progress and record their ideas with short sound bites between the periods of activity. It gave many of the students a chance to "take stock", re-focus and prepare for the next step in the process.

Modelling



Before moving onto the modelling and the second morning it was really good to have an opportunity to share ideas on a wider scale. “Bigging up” students by producing and displaying post-it notes of “best ideas”, “wackiest ideas” etc brought everyone together as a group and gave a fun and lively start to the second session. Most students realised quite quickly that the main purpose of the modelling was to further develop and clarify their designing rather than attempting to produce a real product. Several took this a stage further and produced more than one model- one to show the working of any mechanisms incorporated into the design (often to a larger scale), and another to show “the real thing” at an appropriate scale, looking as realistic as possible.

To say that the modelling is frantic to start with would be not too far from the truth. But as I stated before, being properly prepared is half the battle won. It’s during this time that the demands on the staff are greatest. I would recommend having a “tame technician” on hand to help provide materials... I don’t know why but every time I have done a similar activity there is always a student who needs something that you haven’t put in the collection of materials. It’s good to have someone who can spend a little time running round for them.

In this sort of activity students expect/need instant results with their modelling... no time to waste waiting for glue to dry! So a plentiful supply of glue guns really helps and also some strong double sided tape!

Teacher assessments

Although not a part of the original e-scape project, I volunteered to take part in the assessment of the portfolios. Again Richard and Goldsmiths team provided us with excellent preparation and training to undertake this task. The comparative pairs principle was very straightforward to understand and in practice the online assessment comparisons seemed to go pretty smoothly. The first few pairs of projects were quickly judged and the winners and losers decided. However at this stage I started to struggle in my mind with what I was really trying to assess. It was easy to be “taken in” by a very complete project or a project with a well-finished outcome and not notice that there had in fact been very little creative or innovative thinking going on throughout the portfolio. It became a real challenge to put aside my desire to value the wrong sort of project and start looking for projects that had evidence of “having, growing, or proving design ideas”. Re-reading the performance descriptors and a few emails to Richard probably put me back on track again.

I think that it may well take some considerable time for teachers to become comfortable assessing design innovation. For quite a few years we have encourage students to work methodically through an artificial design process ticking the boxes necessary to satisfy exam criteria along the way. I think it will be us as teachers and assessors who will have the difficulty of changing our mind-set as students will lead the way. I don’t think that this is in any way fault of e-scape, to move Design and Technology forwards we must be prepared to change our values.

There is one thing I am absolutely certain about, I would much rather make 100 pairs judgements than have to mark in detail 20 conventional portfolios. The opportunity to see what other schools produced during the e-scape project really

how to shape my thinking about creativity and innovation in designing, ...some excellent professional development!

Student assessments

After completing the teacher assessments Richard asked if it would be possible to pilot a unique part of the research phase of the e-scape project. The students who had produced e-portfolios at Clevedon School were asked if they like to take part in the assessment process. Even though it meant coming into school on what would have been a day off for them, the majority agreed. The students were shown paper-based examples of previous phases of the e-scape project and discussed the judging criteria. They were shown how by making simple comparative pairs judgements a rank order for the projects could be ascertained. After being introduced to the assessment website the students started to decide on the winners and losers in each pair that was presented to them. It was unbelievable how quickly all of the students manage to get to grips with the assessment process. And listening in to the conversations that were going on during the judgments it was apparent that they didn't have many of the hang-ups that I as a teacher have experienced. Without a doubt they were able to spot the creative and innovative thinking in the design work and were rarely taken in by "pretty" or "content free" products. Although I have not been able to see any comparison to the assessments done by teachers I do have the feeling that they did every bit as well as us ... if not better.

Summary

I suppose the most testing question would be, "would I do it again?" ... to which I can only answer "YES PLEASE!"
Do I see e-scape as an effective way of assessing creativity and innovation in design work? ... Without a doubt!
If Design and Technology is to continue to appeal to students, examiners, universities and employers it has to move on. To keep its credibility D&T has to look to challenging students to not only produce products using tried and tested processes but to show flair and imagination. It's time to start to think about new ways of gathering evidence, assessing what really matters in D&T and to move on. For me e-scape has done this.

Dave White
Subject Coordinator for D&T, Clevedon School.
D&T AST, North Somerset.

15. planning for e-scape phase 4

There are several strands of work planned for phase 4 of e-scape, but the most central is the transition from a research-led operation in TERU to an Awarding Body-led operation that leads to real awards as part of GCSE and Diploma qualifications. Discussions are currently underway involving Becta, Ofqual, Awarding Bodies, QCA and TERU concerning a multitude of issues that need to be resolved in order for this to move ahead smoothly.

The broad plan is that phase 4 will be launched with the new school year in Sept 2009. Schools undertaking some qualifications will have the option of having their candidates taking particular modules of the qualification in e-scape mode. The qualifications that we are currently working on with Awarding Bodies are GCSE Product Design and the Diploma in Creative and Media.

In partnership with each Awarding Body, a group of schools / centres will be agreed and will comprise an Awarding Body pilot programme that will run for the normal 2 years of the GCSE / Diploma course. During that time, agreed modules of work will be undertaken and assessed using e-scape methodology resulting in web-portfolios. It is not necessary for the assessment to be undertaken through a 'pairs' process, but the pilot status of the programme allows us to undertake normal marking and pairs judgements in parallel – to explore the strengths and weaknesses of each.

key features for escape-on-a-stick (esoas)

The central purpose of esoas is that it enables us to run escape activities in classrooms without providing dedicated sets of hardware for the learners to work on. It enables teachers to make use of any of the digital equipment they normally have access to – and for learners to do the activity on those devices through the mediation of the e-scape USB stick.

The system will be device agnostic; working across a wide range of devices and operating systems. No installation will be required – since the main user interface and all the associated evidence-capture tools will run as flash applications on the stick. The only requirement is for the user to have a web-browser and the associated flash plug-in.

The system will also therefore be portable. If the user runs the software from the USB stick they can take their work with them and use the same stick with their saved data on any other machine which is compatible with the escape system. If the session is being run on an internet based server, the student can either use the client available on the stick or can run the session directly from the browser using the web-based client.

In the phase 3 work just completed, the communication mechanism between the client and the server was limited to a 10 second polling routine – all messages were written to the server to be reviewed every 10 seconds by the client. In the new system the communication will be direct and instant – messages are sent directly between the device and the server and vice versa. Messages from the server can also be targeted to individual machines if that is required. The system incorporates instant backup since the changes to the communication layer are also present in the file saving routine. Files from the client are backed up to the server at the point of being saved. This process happens silently in the background.

In the phase 3 work it was not possible for learners to view their portfolios as they were under construction. It will be possible for learners and teachers to monitor their emerging portfolio in the esoas system. Teachers operating an e-scape exam will be able instantly to see when a student has uploaded a file and the system will shows a percentage indicating what proportion of the expected files have yet to be uploaded. Moreover, because the e-scape client is now running in a web-browser, the student has permanent access to their live portfolio. This enables them to review their work at any time just by clicking on the portfolios button. This global view can also be utilised

by the teacher through the EMS to provide formative feedback, even whilst the session is running. See Appendix 31 for the full design specification of the new system.

16. exploring the 'backwash' effect

We have become aware of the possibility that as e-scape approaches become more commonplace through phase 4, we might expect to see teachers adapting their normal practice to maximise the performance of their learner groups. This has been described as the backwash effect of examinations into the curriculum.

Such backwash can be positive or negative. If an assessment process involves and seeks to assess pedagogically desirable qualities (e.g. collaborative performance; articulate presentations; creative idea-modelling), then the most likely backwash effects will be that teachers will seek to find ways of enabling their learner group to get better at these things. On the face of it this would appear to be a good thing – since we would want learners to be better at these qualities. In short, a good assessment process *ought* to generate positive backwash effects into the curriculum.

We decided to make some initial explorations into this issue with a group of the teachers we have worked with in phase 3. Working with teachers in both d&t and science we convened a one-day meeting in TERU and debated the curriculum effects that they would expect to see as the result of wide scale implementation of e-scape approaches.

Both in d&t and in science, the teachers view was that the backwash would be positive. They cited examples from the e-scape activities that they felt teachers would be likely to focus on in normal curriculum time:

telling stories - by which they meant learners getting better at telling the story of their developing work – communicating their intention and process. Science in particular is very bad at this currently – and learners talking through their science (even if in preparation for recording a voice file) would be beneficial.

managing time – this is something that - during lesson times - teachers do not always manage effectively, either leaving too much space or overtly micro-managing. In e-scape, the system helps teachers to focus on *procedural* management, providing a range of time-management interventions. In this way blocks of investigative time (in science) and modelling time (in d&t) can be set up to be entirely the responsibility of learners to organise for themselves. The teacher/administrator is unable to intervene. Teachers and learners getting familiar with this responsibility should improve their understanding of the importance of time management and thereby improve their performance.

Immediacy and consequence - e-scape portfolios build and emerge in real-time and learners have to cope with the consequences of the decisions they make as they go along. This reality-check is in stark contrast to the normal slow evolution of (e.g.)

design project work where teachers (largely) take responsibility for ensuring that everything is OK. Progressively transferring this responsibility to the learner is one consequence of the e-scape style of performance assessment … and a very desirable one.

selecting appropriate tools – the e-scape test requires operation of photo / text / draw / voice / video tools. These are specified into the structure of the activity, but in later iterations are likely to enable learners to choose the tool that is most appropriate for the task in hand. Giving responsibility to learners to select the best media for telling their story and developing their portfolio was thought to be a very positive side-benefit of the approach

coping with unexpected / wrong outcomes – this was noted particularly in science. Where experiments 'go wrong' there is much potential learning in explaining why / how it arose. This is another feature of 'talking' science – and if the e-scape approach was developed into curriculum it would encourage teachers to empower learners not always to be dependent on things working to plan. It places a premium on being able to understand and explain what happened and if teachers 'coach' these qualities it could significantly improve learners performance.

teachers not in control - both in d&t and science it is normal for teachers to tightly micro-manage the classroom / workshop process. In e-scape this was not possible because of the script that controlled teacher interventions. After some nervousness on the part of teachers at their lack of control, they saw real learning benefits from it as learners were required to demonstrate greater autonomy. Getting students used to this autonomy – through increasingly de-structured activities – could be very beneficial for learners.

seeing how the evidence works - in the Clevedon trial we asked learners to judge their own and others' work in a pairs process. The most common response from them at the end of the exercise was "why couldn't we do this **before** the test – so that we could understand what our work needs to look like to be effective". In short, taking part in the pairs assessment showed the group what kinds of performance were effective in the portfolio. Clear, succinct statements in the voice files; photos in focus and highlighting the critical parts of the product etc. Having seen the importance of this for assessment, they would have made sure that their performance was effective.

These and many other matters were raised by the teachers during the course of the day and it is interesting that all the positive backwash elements that they identified are in relation to ***procedural qualities*** of learning. There is nothing here about content; the knowledge and skills of science or d&t. All the issues they have raised are about broader processes involved in making an effective performance – be it in science, d&t or anything else. We were delighted at the teachers' analysis, since it is exactly these qualities that we have sought to incorporate in the e-scape framework from the start.

We do not pretend that this one-day discussion - in which we asked teachers to ***predict*** what they might do when e-scape becomes a more common experience in schools – can do more than indicate likely areas of backwash. Nevertheless it was a very revealing discussion and one that suggests that e-scape backwash might largely be positive. More detailed and explicit treatment of the backwash question will be possible through the phase 4, Awarding Body pilot (see above) and we recommend that it becomes an explicit research question to be dealt with there.

17. other e-scape developments

In addition to the phase 4 Awarding Body pilot programme, a number of other activities are underway and will be developed over the next months and years.

e-scape as assessment for learning

In Scotland, a project is underway to explore the formative, classroom potential of e-scape methodology at the transition years from primary to secondary schools. The project is based in Edinburgh University and is running in collaboration with Becta, Learning and Teaching Scotland (LTS) and the Scottish Qualifications Authority (SQA).

The project "Assessment is for Learning through digital technologies" is directed by Susan McLaren, and has the following aim.

> To develop and extend an adaptation of the 'e-scape' project that captures pupil performance, thinking and creativity, encompassing formative assessment strategies as learners in the transition years work on design challenges.

There are four specific objectives of the project:

- To trial and appraise the opportunities 'e-scape' offers for supporting Assessment FOR, OF and AS Learning.
- To explore the transferability of e-scape technologies and its operability on a range of alternative systems and devices.
- To capture pupil performance in real time scenarios as a means for recognising achievement and developing performance in a wider sense.
- To provide proof of the 'e-scape' concept in the learning and assessing process and how it may affect iterative approaches to creative designing.

There are four schools in the project – and they are currently trialling the e-scape on a stick technology and developing activities with the authoring tool. For the rest of this academic year, learner groups will be undertaking these e-scape activities that have been designed by their teachers. Formative assessment tools are under development with the teacher group and are being built into the e-scape system.

The project is due to report in December 2009.

e-scape towards professional accreditation

A project is under development to explore the potential of e-scape methodologies for building professional accreditation portfolios. In association with TDA and the Geographical Association the plan is that a group of PGCE geography students at the Institute of Education in London will be the focus of this project.

The main purpose of the project is to explore e-scape methodologies for the periods during which students are out in schools and (at least somewhat) disconnected from their tutors and fellow students. Prior to the school placements, a series of e-scape

activities will be developed that the students can undertake while they are out in schools as part of their professional development as teachers. These will be built into e-scape web-portfolios and thereby can be shared with tutors and fellow students. The purpose is principally formative – supporting their development as professional teachers – but equally the portfolios can build to provide evidence of their growing professionalism. Such evidence could subsequently be used as part of the summative accreditation process.

In parallel with this development, Goldsmiths design-education students will be offered the same opportunities – building e-scape portfolios as a development from the 'blog' system that is currently in place for trainee teachers during their school experience.

e-scape in Australia

In Western Australia the Curriculum Council has particular problems with the management of schools assessments. Specifically, they are now required to conduct examinations for all school-leaving certification courses – including those with practical components. This has placed the external examination process under considerable strain – bearing in mind the distances involved in servicing remote communities. They are therefore exploring a number of electronic solutions to this problem.

TERU is collaborating with the Curriculum Council and the Science and Technology Research Centre at Edith Cowan University. We are partners with them in a research grant from the Australian Research Council. The project is directed by Dr Paul Newhouse and entitled

> " Applying digital forms for performance-based assessment in Western Australian post-compulsory courses"
>
> The problem being addressed is the need to provide students with assessment opportunities in new courses that are on one hand, authentic where many intended learning outcomes do not lend themselves to being assessed using pen and paper over a 3 hour period, while at the same time being able to be reliably and manageably assessed by external examiners. That is, the external assessment for a course needs to accurately and reliably assess achievement without a huge increase in the cost of assessment. Therefore the main research question is:
>
> *How are digitally based forms of assessment most effectively used to support the assessment of student performances for relevant courses?*

This three year project is now in its second year and pilot testing will be undertaken in schools in Sept / Oct using – in part – e-scape methodologies. The creation of web-portfolios is possible even in remote communities since the vast majority of them have either school or town satellite/internet connectivity. The Curriculum Council is also keen to explore the pairs assessment process as part of this research.

references

Calvez J.P., Wyche, A., & Edmundson, C. (1993). Embedded real-time systems. New York: John Wiley & Sons.

Hillary N (2005) Measuring performance for real-time systems Document Number: GRNTEEPFRMNCWP 1-800-521-6274, 480-768-2130 Freescale Semiconductor Literature Distribution Center P.O. Box 5405 Denver, Colorado 80217

Kimbell, R., Stables, K., Wheeler, A.D., Wozniak, A.V. & Kelly, A.V., 1991 The Assessment of Performance in Design and Technology: the final report of the APU Design & Technology project, School Examinations and Assessment Council. London, UK 288 pp D/010/B/91

Kimbell, R., Stables, K. & Green, R., 1996 Understanding Practice in Design & Technology Open University Press, Buckingham, UK 126 pp ISBN 0 335 19555 5

Kimbell, R., Miller, S., Bain, J., Wright, R., Wheeler, T., Stables, K., 2004 Assessing Design Innovation: a research and development project for the Department for Education & Skills (DfES) and the Qualifications and Curriculum Authority (QCA), TERU, Goldsmiths College, London UK pp 70

Kimbell RA 2005 "Assessing Design Innovation" ch 1 (pp 17-36) in Technology Education in New Perspectives: Research, assessment and curriculum development Lindstrom L (ed) Stockholm Library of Curriculum Studies Vol 14. Stockholm Institute of Education Press (HLS Forlag) ISBN: 91.7656.591.2

Kimbell RA, Wheeler A, Sheppard T, Brown-Martin G, Perry D, Hall P, Wharf W, Mller S, Potter J 2005 e-scape portfolio assessment (e-solutions for creative assessment in portfolio environments) phase 1 report pp 68 TERU Goldsmiths University of London

Kimbell RA 2007 [i] "Assessment of Design and Technology in the U.K.: International Approaches to Assessment" ch 10 pp 181-203 in Assessment of Technology Education. The 56th Yearbook of the Council for Technology Teacher Education (CTTE) Hoepfl M, and Lindstrom M (eds). McGraw Hill Glencoe. New York New York, Columbus Ohio, Chicago Illinois, Woodlands Hill California ISBN: 0.07.877770.4

Kimbell RA (2007 [ii]) "e-assessment in project e-scape" p66-77 in the Special Edition on Assessment of Design & Technology Education: An International Journal Vol 12 No 2 Design & Technology Association Wellesbourne Warwickshire UK. ISSN: 1360.1431

Kimbell, RA Wheeler A, Miller S, and Pollitt A. 2007 e-scape portfolio assessment (e-solutions for creative assessment in portfolio environments) phase 2 report pp 100 TERU Goldsmiths University of London ISBN 978-1-904158-79-0

Kimbell RA (2008 [i]) "Project e-scape: a web-based approach to design and technology learning and assessment" ch 12 (pp219-241) in The episteme Reviews: Research Trends in Science, Technology and Mathematics Education. Eds Choksi B and Natarajan C,. Macmillan India Ltd. New Delhi. ISBN 10:0230-63443-5 ISBN 13: 978-0230-63443-5

Kimbell RA (2008 [ii]) "Design Performance: DigitaL Tools: Research Processes" in Middleton H (eds). Research Methods for Technology Education, Rotterdam Sense Publishers

Kimbell RA and Pollitt A (2008) "Coursework assessment in high stakes examinations: authenticity, creativity, reliability" Third international Rasch measurement conference. Perth: Western Australia: 22nd-24th Jan 2008

Kimbell, R. & Stables, K., (2008) 'Designerly thinking beyond design' in Borg., K. & Lindberg, V., Kunskap, lärande och bedömning i estetiska och praktiska ämnen (Knowledge, learning and assessment in aesthetical and manual school subjects) Stockholm University Press, pp. 188-198

Mohd Adnan, M. (1993). The role of performance assessment in real-time system management. Journal of Irrigation and Drainage Systems, 7(4), 224-236.

Pollitt A and Crisp V. (2004) "Could Comparative Judgements Of Script Quality Replace Traditional Marking And Improve The Validity Of Exam Questions?" A paper presented at the British Educational Research Association Annual Conference, UMIST, Manchester, September 2004.

Stables, K. (2008). Embedding digital tools in creativity activity: supporting and assessing the development of creativity'. In J.-C. Hong & Y.-F. Pan (Eds.), International Conference on Creativity Development (pp. 33-52). Taipei: National Taiwan Normal University

Stables, K. & Kimbell, R., (2000)‘ The unpickled portfolio: pioneering performance assessment in design and technology’ D&T International Millennium Conference: Learning from Experience: Modelling new futures, DATA, Wellesbourne, UK pp 195-202 ISBN 1 898788 48 0 (refereed)

Stables, K. & Kimbell, R. (2006) Unorthodox methodologies: Approaches to understanding design and technology, in de Vries, M. J. & Mottier, I., (eds) International handbook of technology education: reviewing the past twenty years (pp. 313-330) Roterdam, Sense Publishers

Stables, K., & Kimbell, R., (2007), ‘Evidence through the looking glass: Developing performance and assessing capability’ in Taxén, L., (ed) The 13th International Conference on Thinking, June 17-21, 2007. Norrköping, Sweden (Linköping Electronic Conference Proceedings, No. 21) Linköping University Electronic Press, Linköping, Sweden, http://www.ep.liu.se/ecp/021/vol1/ (pp 175 – 182)

Thurstone, LL. (1927) A law of Comparative judgement. Psychological Review, 34, 273-286

e-scape phase 3 Final Report (Mar 2009)

appendices