Evolving project e-scape for national assessment

Richard Kimbell

Published online: 22 December 2011
© Springer Science+Business Media B.V. 2011

Abstract In the opening paper in this Special Edition I outlined the major issues that led to the establishment of *project e-scape*. The project was intended to develop systems and approaches that enabled learners to build real-time web-based portfolios of their performance (initially) in design & technology and additionally to build systems and approaches to facilitate the web-based assessment of those portfolios. The project was commissioned by the Qualifications and Curriculum Authority (QCA) with additional 'buy-in' from Awarding Bodies—who were seen by QCA as the leading beneficiaries of a successful project. The project was designed in three phases. I have outlined—in the Introduction to this Special Edition—the early exploratory work that we undertook within phase 1, the aim of which was to prove the viability of the concept. This was achieved, and QCA then commissioned phase 2 with a brief to build a working prototype system and run it through a national pilot-testing programme in 2006. Age 15 was the target age-group, aligning as closely as we could with the Awarding Body requirements for the General Certificate of Secondary Education (GCSE) that runs with age 16 learners. The successes of the phase 2 prototype—both as classroom activity and as reliable assessment—led QCA and Becta (the body responsible for funding ICT developments in schools) to commission phase 3 in which we explored the potential of the e-scape system for wider application. Specifically, we were required to demonstrate the *transferability* of the system to other curriculum areas beyond design & technology, and the *scalability* of the system if it were to be used for national assessment purposes, with hundreds of thousands of candidates. In this paper, I outline the approach that we adopted through the e-scape research; describe the major elements of the work both in terms of classroom/curriculum practice and in terms of new approaches to assessment; and analyse some of the key issues that arise from it.

Keywords Web-portfolio · Performance assessment · Comparative judgement · Transferability · Scalability

R. Kimbell (✉)
Goldsmiths University of London, London, UK
e-mail: r.kimbell@gold.ac.uk

Context

In design & technology we have a proud record of coursework-based assessment in which learners project-work is recognised as contributing towards the assessment of their performance. In the Guide for Teachers that was published in 1986 by the Secondary Examinations Council (SEC) at the launch of the General Certificate of Secondary Education (GCSE), it was described in the following terms:

> It has been argued throughout this guide that the exercise of design ability is only comprehensively possible through designing and making in response to perceived need. It is therefore through coursework that a true picture of student development will be gleaned
>
> (SEC 1986 p 37)

However, I have described in the Introductory paper (ppp) how as pressures on schools increased (to achieve good GCSE pass rates), the ability of coursework to operate as an unencumbered diagnostic indicator of learners' capability became increasingly compromised. The essence of the problem—as we saw it—lay in the second-hand nature of learners' design portfolios; essentially reconstructing a tidied-up view of the design development process in order to score as many assessment points as possible. We were seeking to redress this position by making e-scape portfolios real-time, un-editable, direct records of everything that the learner did through the life of the project. Moreover, we sought to create a system that placed no significant burden on learners as their portfolio was building. It should be automatically uploading what they were doing with almost no prompting or intervention by the learner. The creation of the learner's folio happens in the background. We used the expression "the trace left behind [by the learners activity]" to describe this core feature of the portfolios.

So, there was always a powerful pedagogic motive underlying e-scape; a motive that sought to free learners from the burdens of artificial story-telling and allow them just to get on with their designing. This is not to say that learners do not tell a story, for of course they do, by virtue of the rich range of evidence captured in the portfolio. And, informed by our work for the Assessment of Performance Unit (Kimbell et al. 1991), we knew that we should ideally have portfolios that allowed learners to move naturally between phases of *active* designing and more *reflective* review of their work. In a previous project 'Assessing Design Innovation' (Kimbell et al. 2004) we had succeeded in creating a paper-based version of this model of portfolio. The positive responses of learners to it encouraged us to believe that we had indeed created something that was beneficial to them and their designing. It was then adopted by the Oxford, Cambridge and RSA (OCR) examination board and re-badged as the 'Innovation Challenge' one of the examination papers in their Product Design qualification. This adoption was tangible evidence that we had also created an authentic assessment tool that enabled examiners to look in on, and make judgements about, learners' designing.

The e-scape project sought to take a major further step; to create these portfolios digitally, thereby opening up many new forms of data capture for learners. These included *sound bites* for their immediate thoughts and reflection on how they were getting on; *video snippets* of working prototypes; and much more. In phase 1 of e-scape we had explored a range of peripheral digital technologies that were available at the time (i.e NOT desktop computers but handheld devices that could be used freely in workshops). By the end of phase 1, we had done enough to persuade the Qualifications and Curriculum Authority (QCA) to commission phase 2 of e-scape; the construction of a working prototype system to allow all learners’ work to be tracked and logged in a website for subsequent assessment by Awarding Bodies and to test it, in the summer of 2006, through a national trial with year 10 (age 15) learners.

Research questions shaping the project

The construction of the technical system was inevitably guided principally by *technical* research questions, concerning for example;

- the connectivity between hand-held devices in the classroom and web-spaces
- the possibility of pre-defining this web-space so as to construct a virtual portfolio
- the security of access to this virtual portfolio through user-names/passwords
- the robustness of the system
- the management and maintenance of the system

However, the process of developing the prototype was also informed by pedagogic, manageability and functional assessment research questions, for example:

Pedagogic How will the construction and appearance of the virtual portfolio impact upon the questions and sub-activities that need to be built into the activity ? How is the designing activity changed (for learners and teachers) by the system? What backwash effects into the curriculum might teachers anticipate in relation to normal studio/workshop practices?

Manageability how often will the digital devices need to be synchronised to the web-space ? How long does the process take and can a class of (say) 24 learners manage this process simultaneously? How do-able is the digital activity in normal studios/workshops? How much cpd/training do teachers need to prepare for this mode of assessment?

Functional how does the assessment process change when viewing the virtual portfolio in the web-site as opposed to real paper-based portfolios ? The prototype was designed in association with two technology partners Handheld Learning (HHL) and TAG Developments. HHL were specialists in the use of peripheral digital tools and specifically PDAs (personal digital assistants). 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 Developments had 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. The details of this technical development are presented in the paper by Derrick (later in this journal).

Whilst the technology partners were working on the technical system, in the Technology Education Research Unit (TERU) at Goldsmiths we worked on the classroom activity and the protocols that would render it manageable as a design activity worked largely through digital means.

The assessment task and supporting resources

The design task we developed for the phase 2 trials was entitled ‘the pill organiser’ and inspired by the problems we sometimes experience in having (and remembering to take) the right pill at the right time. It was a product design activity to develop a container/dispenser for the pills. Learners had to identify a user group (maybe a 6 year old on a school trip, an active-sports enthusiast or an elderly lady living alone) and think about all the issues involved in the design of a container/dispenser for their pills:

how many pills?
taken how often?
how to remember to take them?
how to keep them secure?
how to make the container/dispenser desirable? etc.

Learners' progress through the activity was facilitated in several ways.

Each student had a working booklet that was used for drawings and other activities requiring more space than was available on the PDA screen.

Each group of 3 students had a *handling collection* of idea-objects. These were readily available inexpensive products that were intended to prompt thoughts about how objects can *contain* or *dispense*.

The handling collections also included a set of 'client' or 'user' cards that profiled particular users and their pill requirements. At a point in the activity students could choose one, or make up a different user altogether.

The group as a whole was also provided with a central modelling kit, including all sorts of soft-modelling materials (paper, card, fabrics, dowel-rod, modelling foam, plasticene etc.) and a range of related constructional materials (tapes, springs, wire, nuts & bolts, pipe-cleaners, cotton reels etc.)

This modelling kit had been the subject of considerable research on our part. It is interesting how the availability of particular resources influences the emerging designs. At one trial, despite some interesting ideas emerging, there was a 'boxy' feel to much of the work. When (in later trials) we supplied more fabrics and (particularly) some plasticene, the variety of responses blossomed. We concluded that:

sheet materials (paper/card) best enable '*boxy'* forms
strip materials (dowel rod/straws/wire) best enable *skeletal* forms
fluid materials (plasticene/clay) best enable *organic* forms
and also that
textile materials (fabrics) often link to and operate across these types

It is as hard to model a mushroom out of rods as it is to model a square container out of plasticene.

The overall choreography of the activity was managed through an administrator 'script' There were several reasons for this. Some were to do with the equity demands of assessment across different schools (making sure the activity was run in a similar way everywhere) but the primary reason was because of the pedagogic concerns about driving the activity forward. We had broken the 6 h activity into (approximately) 24 sub-tasks that were designed to build upon each other. As an example, the task opens with the following sequence:

explore the handling collection and discuss the objects and how they look/work (10 min)
put down (on the pda drawing screen) your first thoughts and ideas about a pill container/dispenser (7 min)

Learners' work is then swopped around the group so that A sees the ideas from B, B sees the ideas from C, C sees the ideas from A. This happens instantly, screen to screen.

look at your team-mates work and (if it was your own) what would you do next to enrich it? (5 min)
work is swopped again to see the other team-mate (5 min)
work is returned to the originator – with all the additions and comments of team-mates.

After just the first 20 min of activity (through 5 steps) learners took leaps of faith, got support from their team-mates and consolidated a starting point. Interestingly we found that it was best to introduce the client/user cards AFTER this initial exploration cycle. As Derrick describes (see later in this volume), from his (technical) perspective these early collaborative stages of the task provided some serious technical challenges.

Through a series of trials we optimised the focus, the timing and the sequence of these sub-tasks to capitalise on the iterative active/reflective process that we knew gave most learners the best opportunity to demonstrate their capability (for the details of this choreography see Kimbell et al. 1991, 2004)

We undertook four task/script trials; first with a Goldsmiths teacher-education student group and subsequently with 3 year 10 groups in schools in the London area. Year 11 groups (the real GCSE year) were invariably not released by schools for such research as it is thought too close to the crunch time of the real examination. It is worth noting however that the task was equally valid for all students; from post-graduates to primary school children. So long as the language level enables the task to be understood by the student group, it is sufficiently open to allow anyone to respond. The quality of the output is then determined by the sophistication of the students' imagination and skill. Interestingly some of the primary school pieces of work in the final pilot were judged to be superior to some of the year 10 pieces (see Kimbell et al. 2009 sects 10 & 11).

By the end of the trialling process we were confident that we had a task (and a set of resources) that students could have a good run at, and show us what they could do.

System trials (April 2006)

After approximately a year of development (see article by Derrick) the beta version of the e-scape application was delivered to TERU on 18th April and we did a complete 'walk-through' of the activity from start to finish, partly to check how it worked and partly to check that it did what we wanted it to do.

Thereafter we undertook a series of real 'run through' trials in schools. The first ever attempt to make the system work was a chastening experience. The school, the teachers and the students worked incredibly hard to help us make the system operate. But the first morning was so disjointed (e.g. system crashes, lost data,pda freezes) that we decided to use the second morning just as a re-run of the first morning, but with a different task.

This trial demonstrated that we needed to be far more careful about our management of the opening 20 min, when students were sharing and collaborating in their work. The protocol we had developed was nothing like the normal start of a design project, and this unfamiliarity was compounded by the technical fragility of the system—moving data from student to student as well as from student to server. By the end of the 2nd second morning we had got to the middle of the activity, and at that point we showed the students what they had done. We projected their work back for them and they were not only fascinated by it, but made the point that it really would have helped to see (at the training session) this 'big picture' of how it was all supposed to work.

The next round of trials was in Cornwall, following a week of modifications to remove the initial glitches. These trials were altogether smoother. The system held up throughout and we only had a small number of pda freezes. This success was due in equal measure to TAGs sorting of glitches in the software, and our re-designing of some of the classroom protocols at the start and end of sub-tasks. Once we knew (from the first trial) where problems might lurk, we could design protocols to overcome them. As a result, in the second trial we progressed well through the challenge of the first 20 min of the activity and on each morning we completed the allotted sections. So by the end of the second morning we had—for the first time—run the 6 h activity right through to the end. As the activity went right through this time, we were dealing with far more of the photography and sound files,both of which caused great interest with the students. Whilst recording the sound files initially caused some embarrassment, they soon learned the process and it rapidly became just a normal part of the activity. (For details of the choreography steps and their rationale see Kimbell 2009).

The final trial focussed closely on the timings necessary for familiarising learners with the draw/text/voice/photo tools in the pda. Our early discussion with developers had suggested that learners would need *weeks* to become familiar with the tools. We subsequently debated leaving the pdas in the school for the preceding *week* so that learners could become familiar with them. For logistic reasons however this was not possible and we were worked with a familiarisation time of 2 h on the day preceding the activity. This general move to shortening the familiarisation time-scale was also based on the reception of the learners themselves to this new technology. They adopted the pdas as an almost natural extension of their mobile phones, having very few difficulties with it. We found that the familiarisation routines had instead to be focussed on the e-scape application interface, which was completely new to them.

Overall, the trials did what they were designed to do. They enabled us to refine the application and taught us how to manage it in the classroom/studio/workshop setting. Throughout the process we were really impressed by the schools, the teachers and the students, without whom we would not have been able to take this important step forward.

By the end of 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 with year 10 students in 12 schools based in three regional locations; Northumberland, Shropshire and the West Midlands; and Cornwall. In each region four schools were identified by the Local Education Authority and examination board consultations.

Training the teacher/administrators

The design of the e-scape system enables it to operate either as a timed formal examination, or as extended and flexible coursework. In the former setting, the task and the resources available to students for tackling it (including materials, time etc.) are controlled by the examination body whereas in more open coursework settings these can be flexibly negotiated. Our work for e-scape phases 2 and 3 was in the context of QCA and awarding bodies and was seen as a GCSE examination.

For reasons of simple equity therefore it was important that the task should be presented in equivalent ways in every school. The digital kit was provided as standard of course, but the teachers had to become familiar with it in advance. Moreover they had to organise their rooms for best effects and provide other resources, e.g. handling collections and modelling kits that were based on a common specification. So there was a training need with all the teachers who were to administer the tasks. With the teachers taking responsibility for managing the task, we were able to adopt more of a *researcher* role during the activities.

Teachers were released from their teaching duties for a one-day training exercise that was organised regionally, and the programme was based on a number of features, including;

- the structure and design of the 6 h activity (why it's like it is)
- the e-scape kit (hard & software + task & resources)
- the experience of the trials (+ showing the website with existing work)
- e-scape training for students in their schools
- setting up the school work-space
- running the activity

At the end of the training sessions, the teachers reported that they understood the system and were confident that they could set up the facilities appropriately in their schools.

Overview of the activity

The arrangements for the start of the 'pill organiser' activity involved a classroom set up with seven working tables each with a group of three learners. The groups had a handling collection to explore at the start of the activity and this contained approximately 15 objects that were intended to stimulate thoughts and ideas about storing and dispensing pills. None of the objects were associated directly with pills but rather with the *concepts* of storing and dispensing. The learners also had their PDA, their booklet, a set of pens for use with the booklet, the user-profile cards, and a set of wooden 'pills' (twice full size) to facilitate modelling. The final element of the room set-up was the table or work-surface set aside for the modelling resources.

Once the task had been established, the early stages of the activity run as described above, with first hazy ideas being swopped between team-mates for some equally early supporting ideas and feedback. When the work returns to the originator, learners are asked to review all the thoughts and ideas that have been offered by their team-mates and to consolidate what they think are the best ideas for moving their product forward.

We then asked learners to reflect on these ideas, specifically in terms of who the *users* will be, and what their specific requirements might be. They did this as text on the PDA, using a keyboard tool and/or the transcriber.

We then introduced learners to the modelling resources. This is an important step since they were typically not familiar with the relatively 'free' notion of material modelling that is central to e-scape. We emphasized that models can be of many kinds, and that early conceptual models might look crude and certainly unfinished but that their value lies in helping us to sort out our ideas. Learners then have a choice of continuing to develop their ideas through sketches and notes in the booklet, or by working with the modelling resources.

Approximately one hour into the activity we asked learners to take a set of photos of their work (booklet or model or both) with their PDA. Having taken them we asked them, to answer two questions; one focussed on *what's going well* with their work, and the other on what is not going so well and that *needs further development*. These questions were posed through the PDA and we ask learners to record their answers as 30 s sound-bites or voice-memos. The rationale for these sound bites at this point lay in the difficulty of interpreting photos in a portfolio. Why were these shots taken? What are they trying to illustrate? By inviting learners to articulate their *thoughts* about the images, it proved much easier to get inside their ideas and understand their designing priorities.

These voice memos were completely new to all the learners involved in the project. It was therefore important that they learned the procedures for doing them and became comfortable with them. The photo-sound-bite routine was repeated 6 times through the course of the activity and learners became skilled at composing and reporting their thoughts. In the early stages however it was important that they could listen to their first attempts and, if they were not content with them, a second attempt erased the previous one.

The activity then proceeded as a series of sketching/modelling phases iterating with the 'pauses for thought' provided by the photo-voice-memo sequence. Typically the modelling became progressively more sophisticated with the ideas behind being subjected to ever-closer scrutiny in the reflection promoted by the sound-bites.

At the end of the first morning and then again towards the end of the activity as a whole, we asked group-members to swop the work around again and advise each other as part of a progress-review process. These were text-based activities in the pda and were taken very seriously by the groups.

Towards the end of the activity, we invited learners to record a 30 s video-clip, essentially explaining the features of their new product and how it worked for the user. It is important to note that the shortness (30 s) of the sound files and video clips was driven by three factors. First to limit the file size and make it easier for the system to move them around; second to focus the learners on creating succinct statements; and third to limit the time for judges to listen to all the files. After a good deal of experimentation in schools we settled on 30 s as an optimum time.

It s important to recognise that after just 5 h of design development and modelling, learners' products were still fairly rudimentary. Many of them were very interesting and innovative – but still rudimentary. We recognised the need for learners to tell us what their final solution would be like when it was completely finished. We labelled this step 'fast-forward' as it was a bit like fast-forwarding a video to the end of the story.

This fast-forward was done as drawing and notes in the booklet, and then photographed into the web-site. It was interesting that some learners struggled to see beyond where they had got to, whilst others were able to envision dramatic developments to transform the existing product idea into a finalised and marketable form (Fig. 1).

The penultimate step of the activity involved learners making a 'pitch' to the managing and marketing directors of a company that might be interested to develop their product

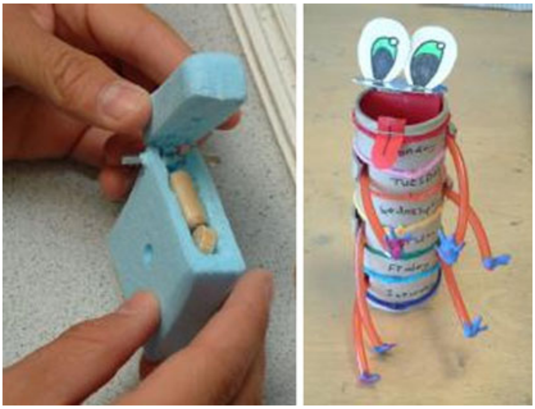


Fig. 1 Two prototype solutions for the pill organiser

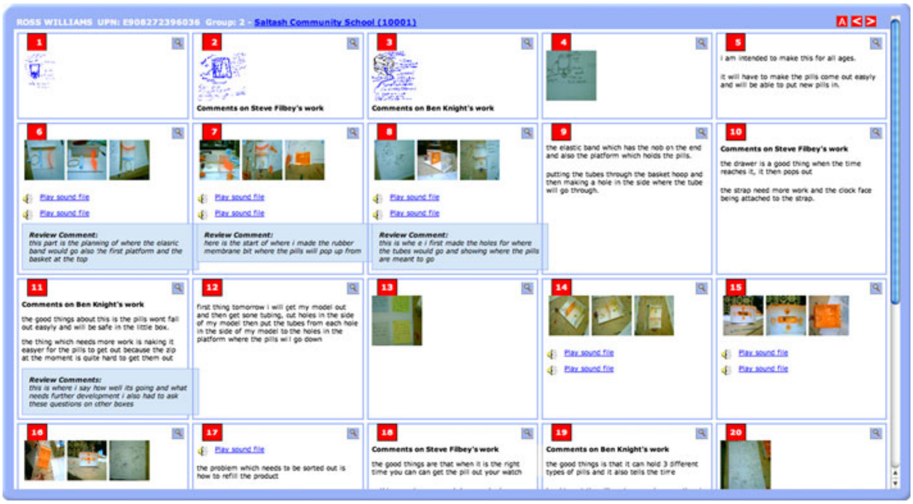
concept into a real product. The scenario is set as if they have 30 s only, in a lift with the directors as they are going up to the 6th floor. They must summarise the originality and potential of their product concept and to persuade the directors to buy the design idea.

Finally, learners review their whole process of design and development, reflecting back (with the benefit of hindsight) on the various steps they have taken. What might they have done differently if they had known then what they know now? Readers should note that as far as possible throughout this whole choreography, the activity iterated between active requirements and reflective requirements, and these were distributed as widely as possible across a range of data collection modes. The iteration process was informed by our APU research and the *Assessing Design Innovation* project, and the distributed data collection issue was informed by our desire to give learners the greatest possible opportunity to develop and explain their thoughts and ideas.

As the work was completed, it was automatically uploaded by the system and located in the website, where each learners' portfolio emerged. The multi-media form of the final portfolio is illustrated below and can be read as a story-board (Fig. 2).

Box 1 is a screen drawing; box 4 is a photo of a bigger paper drawing; box 5 is a text response; box 6 is a series of photos and two sound-files; out of the picture (if you could scroll down) is the video of the learner describing their product and how it works and so on. It is worth noting the pale blue shaded boxes are the learners' comments on their own work, but recorded at the end of the activity as they reflect back on their work.

Each of these boxes of data should be thought of as a thumbnail. If you click on the data in any of the boxes it comes full screen and you can see all the details; examine the photos; hear the learner describing their thoughts in the sound bites; see them describing in the video how it works; hear the 'pitch' and so on. The portfolio view provides an overview; the big picture of everything that is there. But the detail is also there; a click away if you choose to drill down into it.


ROSS WILLIAMS UPN: E908272396036 Group: 2 - Saltash Community School (10001)

1

2
Comments on Steve Filbey's work

3
Comments on Ben Knight's work

4

5
i am intended to make this for all ages.

it will have to make the pills come out easyly and will be able to put new pills in.

6
Review Comment:
this part is the planning of where the elasnic band would go also the first platform and the basket at the top

7
Review Comment:
here is the start of where i made the rubber membrane bit where the pills will pop up from

8
Review Comment:
this is whe e i first made the holes for where the tubes would go and showing where the pills are meant to go

9
the elastic band which has the nob on the end and also the platform which holds the pills.

putting the tubes through the basket hoop and then making a hole in the side where the tube will go through.

10
Comments on Steve Filbey's work

the drawer is a good thing when the time reaches it, it then pops out

the strap need more work and the clock face being attached to the strap.

11
Comments on Ben Knight's work

the good things about this is the pills wont fall out easyly and will be safe in the little box.

the thing which needs more work is naking it easyer for the pills to get out because the zip at the moment is quite hard to get them out

Review Comments:
this is where i say how well its going and what needs further development i also had to ask these questions on other boxes

12
first thing tomorrow i will get my model out and then get sone tubing, cut holes in the side of my model then put the tubes from each hole in the side of my model to the holes in the platform where the pills wi l go down

13

14

15

16

17
the problem which needs to be sorted out is how to refill the product

18
Comments on Steve Filbey's work

the good things are that when it is the right time you can can get the pill out your watch

19
Comments on Ben Knight's work

the good things is that it can hold 3 different types of pills and it also tells the time

20

Fig. 2 The overview of a learners' portfolio, shown as a story-board (box 1-24)

Findings in the background data from learners and teachers

In addition to the performance data arising from the activity, we collected two other sets of evaluative data to compile an authentic account of the e-scape activity and the response of the schools; a learner questionnaire (completed at the end of the activity) and a free-response teacher feedback form (after the event and reflecting on it).

The questionnaire was completed by 256 learners and the full details of the findings are reported in Kimbell et al. 2007. The data confirm the informal impression created in the activity that learners very rapidly got to grips with the device and its associated software, adopting them as merely an extension of the mobile phones that they all use ubiquitously. This is confirmed by learners' response to the statement 'it was easy to learn to use the PDA'. 99% of learners agreed or strongly agreed with the statement. And 98% agreed or strongly agreed with the statement 'it was fun using the PDA'.

The second part of the questionnaire asks about particular features of the PDA and learners reactions to them as part of the activity.

- 96% agreed or strongly agreed that it was easy to follow the activity through the e-scape interface.
- 94% agreed or strongly agreed that it was good for making the photo story lines.
- 92% agreed or strongly agreed that it was good tool for designing.
- 90% agreed or strongly agreed that the text tool was good for explaining ideas.
- 89% agreed or strongly agreed that they were able to show what they could do.

Within all of these sections of the data there is no significant gender variance. The only significant gender effect is observable in response to the statement 'the voice memos were good for explaining my ideas'. 50 boys but only 24 girls *strongly* agreed with this statement, whilst one boy but 14 girls *strongly* disagreed. In response to the voice-memos therefore, at the extremes of the data there is a clear effect that suggests girls are less likely to appreciate it. The less extreme data (agree or disagree with the statement) is gender balanced, and overall 70% of learners agreed or strongly agreed that the voice memos were good for explaining their ideas. We believe that the identifiable gender effect at the extremes is related to the embarrassment/discomfort that some learners felt in talking about their work in the public arena of the activity and the working groups.

Comments were sought from teachers on several issues and the following comments were typical of those returned to us.

Concerning the *task* (pill organiser)

- I liked it because it was a new idea for the children – probably something they had not even considered before (RR)
- The pills task was unfamiliar to our students… I think this helped them approach the problem with fresh minds and with less pre-conceived ideas. (AP)

Concerning the *activity*

- The activity structure works quite well and maintains pace and focus, I have been trying a similar approach with some of my key stage three groups with reasonable success (RR)
- I was very pleased how the children stayed on task even thought they must have been flagging by the end. I think this was due how the task was structured as well as the eagerness to do all things digital. (Nan)

Concerning the *e-scape system*

- I was amazed how quickly the children grasped the technology and were in no way over-awed by it (I shouldn't have been!) (Nan)
- I was particularly impressed with how they used the voice recordings and took them so seriously. I feel this has tremendous potential for capturing their thoughts and feelings about their work as they are doing it. (PD)
- They found the novelty and ease of use of the PDA's a positive motivator (RR) (all extracts from Kimbell et al. 2007 pp 40–43)

Through June and July 2006 we observed as teachers ran the pill activity in 12 schools across the country. Our 13th school was a primary school in Cornwall—since we were keen to see how the year 5 children (10 years old) would cope with the task and the technology. In the event, running it with these children was almost indistinguishable from running it with year 10 (15 year old) learners.

As a result of all the school-based activity we accumulated 250 e-portfolios in the website. The system worked remarkably smoothly and we are grateful for all the support and enthusiasm from teachers and learners.

Making assessments

At the start of phase 2 of e-scape we had assumed that the assessment process would be a continuation of the normal criterion-based, rubric-style assessment process. We knew we could make it screen-based, and hopefully therefore more convenient, but nevertheless we expected the assessment itself to be conventional. But as we grew the digital approach we were gradually introduced to more and more digitally informed experts, and one in particular had a profound effect on what we did for assessment.

Alastair Pollitt had worked with the University of Cambridge Local Examinations Syndicate, and subsequently headed the research section for Cambridge Assessment. He pointed out that all assessment involves a comparison of one thing with another and he introduced us to the scary world of 'comparative judgement' (see the following article by Pollitt).

As Pollitt points out, it is not obviously necessary that exams should be marked. We need to have some way to *judge* students' performances, but there are good reasons to prefer holistic judgments over atomised analytic ones. Pollitt helped us to build a completely different assessment tool that was based on comparative holistic judgement. *Comparative* judgement is *relative*; comparing this to that, It does not require the precision of *absolute* judgement (scoring on an absolute scale) and therefore we are very much better at it.

Imagine yourself in a house, moving from room to room. Some are warmer and some are cooler. You would have no difficulty at all in saying that room X is (relatively) warmer than room Y (assuming that it was). But if I ask you to do that on an absolute scale—what is the Celcius temperature in room X and room Y—then all sorts of error would creep into your judgements. And it is exactly the same with educational assessments. In assessment research projects—over and over again—the thing that teachers are invariably good at is ranking (i.e comparative judgement). It is easier for a teacher to say that child A is a (relatively) better reader than child B, than it is to say that child A is (absolutely) reading at level 4.

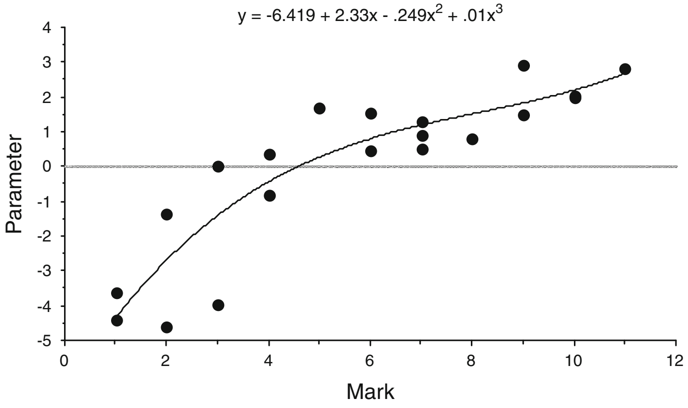
In association with Pollitt and the TAG Developments software team, we built an assessment tool that uses this phenomenon. It presents pairs of portfolios to judges (teachers) and, guided by a set of criteria, asks them to identify which of the two is the stronger piece of work. This simple judgement process is then repeated many times; comparing many portfolios; with many judges. And this results in a rank order emerging.

When we were first introduced to the idea, it seemed too good to be true, so we set up a trial to see whether the comparative-judgement system would replicate a rank order that we had already established. The trial used 20 paper-portfolios/'scripts' from *Assessing Design Innovation* that had been marked in the conventional way on a 12 point scale. The 20 pieces were selected from across that scale. They therefore represented a rank order about which we were confident. Pollitt then divided the scripts into paired groupings for the research team to judge. Each piece was compared to at least 6 other pieces and was judged by at least four judges. There were 6 judges involved in the trial and we each had 40 judgements to make; i.e. a total of 240 pairs to differentiate. With the scripts in the middle of a big table, we all sat around and pulled out the pairs we needed to compare and then returned them to the pile. All we had to say (by reference to the same overall criteria of performance) was which of the two sample portfolios represented the better performance. We had to read the work on both pieces, understand it and make a judgement. To make the process quicker, Pollitt created 'chained' pairs; requiring comparisons of, for example:

Portfolio 3	and	Portfolio 17
17	and	6
6	and	15
15	and	4

… and so on. This ensured that the judges only had to take in one new piece of work to make each paired judgment. The following statistics (based on Rasch analysis) were then prepared by Pollitt from his analysis of the judgements made by all of the judges in the trial.

> To confirm the results of the analysis the scripts' parameters were plotted against the marks previously assigned to them. As expected, there was a strong but non-linear relationship between the parameters and the marks. (Kimbell et al. 2007)



The non-linear relationship between the judging parameter scores and the numerical mark

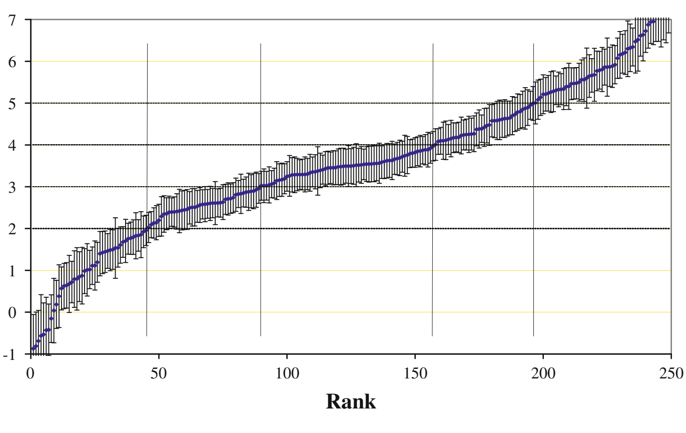
We were very aware of three key measures for the effectiveness of an assessment system. Does it provide *valid* measures of capability? Is the assessment *reliable*, and is it *manageable*? Validity is a measure of the extent to which the resulting number is an authentic way of representing the design & technology capability of the learner. Reliability is a measure of repeatability; do all judges say the same? And manageability is (as the name suggests) about whether the judgements are do-able in a reasonably efficacious way.

We were convinced of the validity of this approach. Using holistic judgements (informed by criteria in the rubric) we were able to make overview judgements that reflect the complex interactions of elements of each learner's work. The reliability statistic was very good because (unlike normal marking) the work is seen in relation to many other pieces and by many judges. It is therefore, by definition, a 'repeatable' judgement. The problem was manageability. When judge 1 wanted to see piece portfolios 3 and 17, judge 5 was already looking at portfolio17 and comparing it to 10. The logistic difficulties of *not* having the 20 pieces available to all judges all the time proved very time-wasting.

But when all the portfolios are in a web-site, they are all available, all the time, to any judge who logs into the system, and from wherever they log in. Up to now in the assessment world, the comparative judgement approach to assessment has been *only* a research tool because of the logistic problems of making it anything bigger. The *e-scape* system (with all the work to be assessed held in digital form) provided, for the first time, the opportunity for comparative judgement to be used simultaneously by as many judges as necessary.

Following the success of this trial, we worked with Pollitt and the software development team to build an automated version of the comparative pairs process (see articles by Pollitt and Derrick). The 'pairs engine' emerged in the summer of 2006, ready for the first ever large-scale assessment of performance using this approach. We recruited and trained a team of judges who made all their judgements in October 2006. Using Rasch analysis, the mass of wins and losses for individual portfolios were automatically 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. Distributed in between are those that win some and lose some.

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.



In the report that Pollitt prepared as a result of this assessment round, there are three nuggets of information to which I would, in particular, draw the attention of readers, quite apart from the performance scale itself.

First the *reliability* of the resulting scale.

> The key figure here is the reliability coefficient of 0.93. This figure allows for unreliability between markers *as well as* for lack of internal consistency within the examination—most traditional reliability coefficients only allow for one of these. Only a few current GCSEs are likely to be as reliable as this if we consider both sources of unreliability. (Pollitt in Kimbell et al. 2007 pp 51–53)

But this reliability is hardly surprising. Each piece of work has been compared with many others, and judgements have been made by many judges. Any idiosyncratic judgements are soon outweighed by the weight of opinion of the team. The process is almost inevitably more reliable than current GCSE practices, where much of the work is assessed by the teacher alone, or at best by the teacher and one external moderator.

Second it is important to note the consistency of the judges. In this comparative pairs approach, the analysis automatically produces a measure of the consensuality of the judging team. The system notes how often, and by how much, one judges' decisions are at variance with other judges and in the end produces a mean score for the whole sample. If I am more than two Standard Deviations from that score, then I am a cause for concern. As Pollitt reported; 'None of the judges failed this test'.

Third, the system also automatically produces data on the consensuality of judgements applied to individual portfolios. Reference to the graphic representation of the 'plot of values' (above) shows some portfolios with much longer standard error 'tails' than others. These are the portfolios over which there was a considerable amount of disagreement within the judging team. So in the process of generating the rank, the system automatically highlights the pieces of work that need closer attention.

These three features: the reliability of the scale, the consensuality measure of judges, and the identification of any portfolios that generate excessive disagreement, are all automatic virtues of the comparative pairs judging process.

(For a fuller account of the judging process from phase 2 see Kimbell 2007a)

Conclusions from phase 2

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

Concerning *technological* challenges The key point here is that we demonstrated how the whole system could be 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 the learners' evidence trail in the web portfolio.

Concerning *pedagogic* challenges The key point here, attested to by teachers and learners alike, is that the iterative active/reflective steps in the choreography, along with the broadly distributed data-collecting modes 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 the splendid isolation of IT suites. In e-scape we showed how the technology could get down to where it was 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. 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 created a system that operates right across this terrain. Learners and teachers reported that it worked well at the performance end, and the out-flow data showed that it produced reliable statistics at the assessment end. The prototype had done enough to demonstrate that it was a functional system for assessment.

The data and outcomes from phase 2 of the project were 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 2007a, b, 2008a, b, Kimbell and Pollitt 2008; Stables and Kimbell 2007; and Stables 2008.

The policy need for phase 3

In phase 2 we had established that the technology for building web-based portfolios from the classroom was quite workable. Moreover we had demonstrated how such portfolios might be assessed with remarkable reliability, again using web-technology. But there were of course limitations to phase 2, arising from the fact that it had all been undertaken in *research* mode. It was exploratory and throughout it was focused on only one curriculum area—design & technology. So despite the obvious successes of the research and development process there was additional work to be undertaken if the approach was to become anything like mainstream. 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.

This step involved two principal extensions of the work.

(1) Concerning specifically the *transferability* of the e-scape system 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 was not sufficient to have the system that can 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.

An overview of phase 3 issues and approaches

Phase 3 of project e-scape ran from Sept 2007 to March 2009 and pivoted around the pilot testing that took place in June and July of 2008. Everything up to that point was as preparation for it, and everything after that point was concerned with the assessment and reporting of performance within it.

Initially we recruited two new subject teams—one for science (directed by Prof Dan Davies at Bath Spa University) and one for geography (directed by Prof David Lambert at the Institute of Education London). Both teams had to develop tasks that were appropriate for e-scape-style performance tasks, trial them as paper-based activities and be ready to run them electronically as soon as the digital system was ready.

Whilst this process was underway the technology team in TAG was building a new interface, the ‘authoring tool’, that enabled teachers to create their own activities within the e-scape system. In phase 2 we had to hard-wire the activity into the system, and once written it could not be changed. But our vision for phase 3 was that teachers could use the new task creation tool to build their own activities and modify them at will in the light of experience in running it.

Meanwhile, the central TERU team set about developing training and resources materials for teachers; helping them to understand and develop the skills and procedures they would need to employ in order to run the system effectively as assessments.

The geography and science work progressed through a series of tasks and trials. Science explored ‘road safety’ as a generic theme for all kinds of science activity whilst the geography team explored a series of typical geography activities—mineral extraction; congestion relief road; rain-forest; town re-generation—and trialled them in numerous schools. The science task emerged as a 3 h (morning) activity and geography as a whole day (5 h) task, part in school and part off site for field-work data collection.

Meanwhile the technology team (using a series of trials run with the central TERU team) refined the software system to beta version and finally to 1.0 release version for the national pilot. This became available in March 2008 allowing the subject teams to ‘load’ their various activities for pre-pilot testing.

The TERU team worked with QCA/Becta/Awarding bodies to identify a range of testing schools for the June pilot. This was arranged through four regional centres: The south-west (north Somerset/Bristol); west midlands (Shropshire/Birmingham); north east (Northumberland); and the south east (London). Having identified schools, we prepared training materials for teachers, conducted the training and arranged final plans for the national pilot. This process gave rise to a debate about the need for two kinds of approach in schools.

As a minimum, it was necessary to know that the system could be used ‘*cold’* by a school, in the sense that they could just be sent the task (downloaded from a web-site), run the activity, and upload the resulting data. But we also believed that the process had the potential to influence teachers’ approach to tasks right across the curriculum. We therefore created a set of ‘*hot*’ schools where the kit would be left for an extended time. In this way we hoped that teachers would absorb the technology into their normal curriculum practice.

After a series of pre-pilot tests, in which the subject teams became familiar with the technology and the protocols necessary to run it effectively in schools, the national pilot was run in schools across the country. The design of the design & technology element of national pilot was aimed at informing the ‘scalability’ question and therefore had a number of variables.

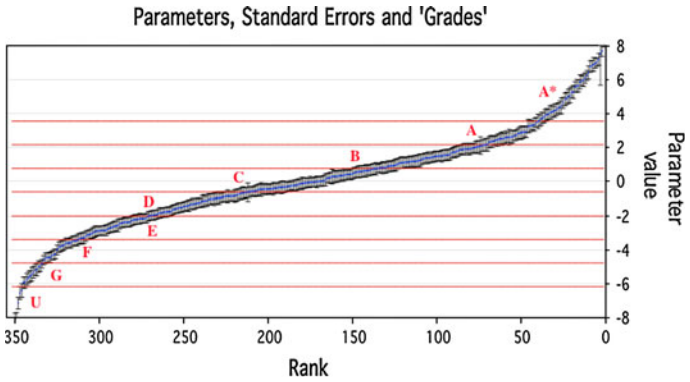
(1) “hot” schools had previous experience of the e-scape system and had the technology for an extended period in which to embed it into their curriculum.
(2) “cold” schools were selected only for national pilot purposes and were therefore running the e-scape task ‘cold’.

These different groups allowed us subsequently 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 team re-developed the ‘pairs engine’; the judging system algorithm and interface so as to be ready for undertaking the judging in Sept/Oct 2008. Meanwhile the TERU team oversaw the national testing of the d&t task, recruited the judging team and prepared training materials for judging training.

The judging process began in September 2008 with the training of the judging team and immediately thereafter the judging itself. The June/July testing had generated 350 d&t portfolios and we had 28 judges. All the teachers in the pilot schools had volunteered to be judges—and we had the TERU team as well as several volunteers from Awarding Bodies who were keen to experience the process. Additionally science and geography had 60 portfolios each (three schools each), and had judging teams of 6.

The judging proved even more easy and reliable than it had been in phase 2. Eased by the relatively simple interface, the judging process became much smoother and quicker, and further improved the reliability. It is worth highlighting Pollitt’s summary observation on this process.


Parameters, Standard Errors and 'Grades'

> 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.
> (Kimbell et al. 2009 pp 62–70)

At the end of the testing and judging process we asked teachers to comment on their experiences of the whole process. The following two comments are typical of the responses of the 28.

> 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)

> Over time, this assessment technique would be likely to enhance the quality of practice of 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)
> (Kimbell et al. 2009 pp 59–62)

The final stages of the project were devoted to data analysis and reporting. In the process we showed how the system operated effectively at the classroom level, enabling learners to create e-portfolios of their performance. Additionally however we showed how the e-scape system could be operated as a national system of assessment that would enable all of the following:

- 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 for 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 completed this full simulation of the e-scape coursework assessment system in a way that enabled Awarding Bodies to identify how they might take it forward into formal GCSE pilot programmes.

Conclusion

To conclude this account of the e-scape development story I would like to draw readers' attention to an important, and often unconsidered, consequence of testing and assessment. I refer to what the literature labels 'the backwash effect' of assessment into the curriculum. As an example, in their study of English as a Foreign Language (EFL) testing, Riaz and Razavipour (2011) conclude that

> the heavy shadow of centralized tests and their strong negative backwash effect have downplayed teachers' agency in favor of the dominant structure. There was clear evidence of the backwash effect of the summative exams on teachers' agency.… The pattern clearly shows the backwash of the summative examination procedure on the classroom teaching and learning activities. (p 138)

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.

At the end of the project, we convened a one-day conference of all of the teachers involved in the project and debated the curriculum effects that they would expect to see as the result of wide scale implementation of e-scape assessment approaches. They listed the following as characteristic of the things that they themselves would make sure they did with their students.

Telling stories helping learners to get better at telling the story of their developing work; articulating (through conversation and recorded sound-files) their intention and process. See for example Roth (2005) concerning students 'talking science' so as to better understand what they are seeing and doing.

> Writing science is only one aspect of *doing* science; talking science may in fact constitute a much larger part of accomplishing the various aspects that characterise science. (p1)

Managing time 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. 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, so 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 progressing satisfactorily. Progressively transferring this responsibility to the learner is one consequence of the e-scape style of performance assessment.

Selecting appropriate tools e-scape activities required the operation of photo/text/draw/voice/video tools. These were specified into the structure of the activity, but in later iterations we enable learners to choose the tool; they decide which is the most appropriate tool 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 (see above). 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.

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. Whilst the teachers administration script might specify that 'you now have 30 min to continue developing your ideas' it deliberately does *not* say how learners might do that. Learners had to decide for themselves whether to draw or model their ideas; whether to concentrate on the mechanism or the appearance; whether to do it real size or double size; etc. etc. 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 learners used to this autonomy—by progressively specifying less and less—could be very beneficial.

Seeing how the evidence works in one school we asked the 15 year old learners to judge their own and others' work in a comparative pairs process. The most common response from them at the end of the exercise was "why didn't you show us 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 out-turn of their activity in the portfolios, and how they worked for assessment, the learners were clear that they would have presented their work differently; more clearly; more persuasively.

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 were in relation to *procedural qualities* of learning. There was nothing here about content; the knowledge and skills of geography or science or d&t. And equally (and very interestingly) there was nothing about the technology. All the issues the teachers raised were 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 was exactly these qualities that we had sought to incorporate in the e-scape framework from the start. As I pointed out at the beginning of this piece, despite the apparent dominance of assessment in our thinking 'there was always a powerful pedagogic motive underlying e-scape'.

References

Kimbell, R. A. (2007a). Assessment of design and technology in the U.K.: International Approaches to Assessment chap 10, Assessment of technology education. In M. Hoepfl, & M. Lindstrom (Eds.), *The 56th Yearbook of the Council for Technology Teacher Education (CTTE)*. McGraw Hill Glencoe. New York New York, Columbus Ohio, Chicago Illinois, Woodlands Hill California.

Kimbell, R. A. (2007b). E-assessment in project e-scape (pp 66–77) in the Special Edition on Assessment of Design & Technology Education: An International Journal Vol 12 No 2 Design & Technology Association Wellesbourne Warwickshire UK.

Kimbell, R. A. (2008a). Project e-scape: A web-based approach to design and technology learning and assessment ch 12, B. Choksi & C. Natarajan (Eds.), *The episteme reviews: Research trends in science, technology and mathematics education* (pp. 219–241). New Delhi: Macmillan India Ltd., ISBN 10:0230-63443-5 ISBN 13: 978-0230-63443-5.

Kimbell, R. A. (2008b). Design performance: DigitaL tools: Research processes. In H. Middleton (Ed.), *Research methods for technology education*. Rotterdam: Sense Publishers.

Kimbell, R. (2009). Performance portfolios: Problems, potentials and policy, chap 42. In A. Jones & M. de Vries (Eds.), *International handbook on research and development in technology education*. Rotterdam, NL: Sense Publishers.

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), Technology Education Research Unit [TERU], Goldsmiths University of London, London UK.

Kimbell, R. A. & Pollitt, A. (2008). Coursework assessment in high stakes examinations: authenticity, creativity, reliability. In *Third international Rasch measurement conference*. Perth: Western Australia: 22nd–24th Jan 2008.

Kimbell, R., Stables, K., Wheeler, A. D., Wozniak, A. V. & Kelly, A. V. (1991). *The assessment of performance in design and technology*, School Examinations and Assessment Council. London, UK HMSO D/010/B/91.

Kimbell, R. A., Wheeler, A., Miller, S., & Pollitt, A. (2007). *E-scape portfolio assessment; phase 2 report technology Education Research Unit [TERU]*. London, UK: Goldsmiths University of London.

Kimbell, R., Wheeler, T., Stables, K., Shepard, T., Pollitt, A., & Whitehouse, G. (2009). *E-scape portfolio assessment: Phase 3 report Technology Education Research Unit [TERU]*. London, UK: Goldsmiths University of London.

Riaz, M., & Razavipour, K. (2011). Agency of EFL teachers under the negative backwash effect of centralized tests. *International Journal of Language Studies (IJLS)*, *5*(2), 138–145.

Roth, W. M. (2005). *Talking science: Language and learning in science classrooms*. Lanham, MD: Rowan & Littlefield.

Secondary Examinations Council [SEC] (1986) *Craft design & technology GCSE: A guide for teachers*. Milton Keynes: Open University Press.

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. (2007). Evidence through the looking glass: Developing performance and assessing capability. In L. Taxén (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.