Technology and the assessment of creative performance

Professor Richard Kimbell
Technology Education Research Unit
Goldsmiths College London
r.kimbell@gold.ac.uk

keynote presentation at
Cambridge Assessment Conference
15th Oct 2007

1. Background

1.1 developments in design & technology

In 1999, a version of the National Curriculum was published, that - for the first time - articulated 'importance' statements, in which the vision of subjects is encapsulated. The Statement for design & technology reads as follows:

> The importance of design and technology
> Design and technology prepares pupils to participate in tomorrow's rapidly changing technologies. They learn to think and intervene creatively to improve quality of life. The subject calls for pupils to become autonomous and creative problem solvers, as individuals and members of a team. They must look for needs, wants and opportunities and respond to them by developing a range of ideas and making products and systems. They combine practical skills with an understanding of aesthetics, social and environmental issues, function and industrial practices. As they do so, they reflect on and evaluate present and past design and technology, its uses and effects. Through design and technology, all pupils can become discriminating and informed users of products, and become innovators (DfEE 1999 p15)

It was noted that the vision statement encapsulates the need for creativity, innovation and teamwork in design & technology and that these qualities were largely missing from the formal modes of assessment that are applied at the point of national examination (GCSE) at age 16. The DFES Strategy Group recommended that research be undertaken 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 GCSE. The Technology Education Research Unit (TERU) at Goldsmiths College was asked to undertake the work and develop a system of assessment that would measure and reward design innovators. The project – *Assessing Design Innovation* - was launched in Jan 2003 and concluded in Dec 2004

The principal outcome of the 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. All the learners' work was structured into an A4 workbook that folded out to become an A2 sheet. The activity was designed to be administered by teachers in ordinary design & technology facilities. The workbooks were carefully designed to unfold throughout the activity, ensuring that students always had sight of the instructions for the sub task they were currently working on and the work they had just completed.

1.2 developments in e-learning

The present government has embarked on a major programme to digitise many of the activities and services it offers, driven by (among other things) the promise of greater control, improved efficiencies, cost savings and better standards of service. This focus on developing new ICT systems straddles many aspects of government from (e.g.) taxation, registration, legislation, communication, health and education. In this education 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.

Beyond government initiatives it is also evident that the use of mobile digital technologies has dramatically swept through the population over the past 5 years, and not least so with young people. One example of this phenomenon is provided by research conducted by MORI for the Nestlé Social Research Programme into the role of mobile phones in young people's lives. Moreover it is not just that they have access to the technology, they also use it; with 9 out of 10 texting at least once a day and over 25% taking photos daily.

Access to:	Year 9-10	Post 16 in full time education
Mobile	95%	99%
Internet	85%	95%
e-mail	70%	90%

(Haste H. 2005)

We are witnessing the growth of 3rd generation computing. Mainframe computer technologies of the 1960s and 70s gradually faded with the emergence of 2nd generation 'desktop' computers. These completely transformed our working relationship with computers, providing us with far greater interactivity, apparently unmediated by the programmers whose services had formerly been essential. We could 'drive' our own 2nd generation computers in the 1980s and 90s . As the technologies shrank, the growth of laptop computers particularly in the final decade of the 20th C did not materially change our relationship to computers. They operated merely as slightly (very slightly) more mobile versions of the desktop. But the new 3rd generation of computers is radically different. PDAs (personal digital assistants) and 'smart' phones are FAR more mobile, are equally powerful, and can now genuinely be regarded as 'back-pocket' computers. As such, they are in the process of transforming – once again - our working relationship with computers. The transition to 3rd generation mobile technologies will be just as dramatic as was the transition from the 1st to the 2nd generation. In the contexts of learning, teaching, curriculum and schools, these transformations will be profound.

1.3 e-learning in design & technology

In schools, design and technology is recognised as an area in which wide and effective use is made of the new digital technologies and statistics from the annual DfES survey of ICT in Schools reflect increasing use and positive effects.

Use of ICT in areas of the curriculum	d&t	English
substantial	62%	19%
some	35%	69%
little/none	3%	12%

Positive effect of ICT in areas of the curriculum	d&t	English
substantial	64%	24%
some	32%	63%
little/none	4%	13%

(DfES 2003)

> Secondary design and technology (D&T) departments continue to make widespread and effective use of ICT in their teaching. (OFSTED 2004)

1.4 the case for e-assessment

In design & technology alone, approx half a million students are assessed annually based on their performance in coursework projects. Students produce a 'portfolio' through which they develop a design solution to a task of their own choosing, and simultaneously tell the story of their development process. Approx 50% of their GCSE marks are allocated on the basis of the quality of their portfolio.

Awarding Bodies responsible for these assessments – particularly at GCSE (age 16) – are increasingly seeking to exploit the power of digital technologies and the importance of e-portfolios within this strategy has been underlined by OFSTED in their recommendation concerning the development of ICT in schools. They make clear that at the school level there is a need to:

> develop electronic portfolios of pupils' work alongside the use of web or intranet-based applications that enable assessed work to be easily accessed by teachers, pupils and parents (OFSTED 2004 [ii])

1.5 digital enhancement of *Assessing Design Innovation*

It was during the development of the activities for *Assessing Design Innovation* that we became aware of the potential for digital enrichment of the activity. Learners increasingly use digital technologies as part of their work in design & technology and these thoughts led us to develop a proposal to QCA/DfES for a digital approach to portfolio assessment. The time was right to explore the e-learning and e-assessment implications of this 3rd generation of computers, and design & technology provided fertile ground in which to undertake this exploration.

2. The structure of project ***e-scape*** [1]

2.1 The brief

As a result of a series of discussions between the DfES, the Qualifications and Curriculum Authority (QCA) and TERU, the following brief emerged

> "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)

[1] ***"e-scape"*** stands for 'e-solutions for creative assessment in portfolio environments'

The work for project ***e-scape*** was divided - broadly - into two areas of concern. The first was with the ways in which digital technologies might be used ***to enhance learners' designing*****.** This was the priority concern of the research team at the outset, since we were determined to ensure that any digital systems introduced into the designing activity should operate as an *enhancement* to the activity - rather than a distraction or a distortion. Accordingly we worked with schools - some of which had been involved in the 'assessing design innovation' project - and explored a range of technologies with learners.

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 ***webspace*** - accessible to the learners themselves, and their teachers, and (ultimately) to examination board assessors.

2.2 findings from the school trials (Jan-May 2005)

These early explorations resulted in a set of findings that we reported to QCA under the following four headings

- technological – concerning hardware, software and data management systems
- pedagogic – concerning the effects and potential in the classroom
- functional – concerning formal assessment requirements (eg reliability)
- manageability – concerning the workability of the system for school-based assessment

These findings – taken together – informed the development of a specification for a prototype system that we planned to build as phase 2 of the ***e-scape*** project. Furthermore, having built the prototype, we planned to run a pilot study in June/July 2006 to test the prototype in a number of schools throughout England

Within this pilot we planned that groups of 16 yr old students would be working on assessment tasks in normal design studios and workshops. In response to set tasks, they would be designing & developing products using PDAs as digital sketchbooks, notebooks, cameras, and voice recorders. The prototype would enable their work automatically and simultaneously to be sent through a wi-fi connection to a secure web-space in which their virtual portfolio would emerge. This virtual portfolio develops through the 6 hr activity and can be viewed alongside their real material modelling of prototypes. Once completed, the work could then be accessed and assessed remotely by Awarding Body moderators.

We should not underestimate the dramatic transformation that this approach holds for design & technology both in terms of pedagogy and assessment. Project ***e-scape*** creates the potential for a fundamental re-think of our traditional practices in design & technology, but the pilot of our prototype system had to operate within existing circumstances. It does not help to frighten the profession too much too soon.

2.3 *e-scape* system components

The escape system operates at several levels:

i. The school workshop/studio activity

The activity (with in-built design task) is self-contained in the sense that the research team take ALL necessary materials and equipment to the school and remove it thereafter. We recognise the demands we make on teachers and schools and we endeavour to make the experience as manageable as possible for teachers and learners.

ii. The pda

The central transformation of ***e-scape*** (moving on from the '*innovating assessment'* project) is that the paper-based work has been replaced by work on hand-held computers – PDAs – in this case the HP3715. The strength of this tool lies in its multi-functionality. It can be used for *drawing*, for *note-writing* (based on handwriting or virtual keyboarding), for taking *photos* and for recording *voice memos*. Moreover the pda has Bluetooth and wi-fi capability, enabling us to link a class-set of them into a local area network. It therefore has interesting potential not only as an assessment tool, but also as a collaborative (group-based) learning tool.

iii. The local area network

This is run from a lap-top computer managed by the teacher/researcher administering the activity. The PDAs all have the ***e-scape*** software loaded into them in the form of 23 linked screens through which the activity progresses. The administrator controls the activity through the laptop. As s/he activates "box 1" on the administrator interface, the signal is sent (via wireless router) to all the PDAs and they all "come live" with box 1. Learners can draw/make notes on the pda, and at the end of the allotted time for box 1 a warning screen pops up prompting learners to 'tap-here' to save their work. This is then sent back through the router and stored in the laptop – as well as in the memory on each individual PDA.

iv. Uploading data

Every 5 minutes throughout the activity, learners students' work (including drawings, notes, photos, and sound files) is automatically transmitted (by wi-fi) to the administrators lap-top and from there into a web-space. We also have a routine to save all the data from a centre on a 500 mb memory-stick attached to the laptop. This "belt & braces" technique is designed to secure the data in case of a problem with the primary system. This primary system involves linking the laptop to an internet-connected computer so that all data is up-loaded directly into a secure web-space.

2.4 the *e-scape* trials

The beta version of the ***e-scape*** application was delivered to TERU on 18th April 2006 and we did a complete 'walk-through' 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. We immediately identified some modifications that we requested prior to formal trials.

The ***e-scape*** prototype was trialled in three schools over a period of about a month and between each trial we undertook revisions both of the software and of the pedagogic structure of the activity.

The 1st trial was somewhat chaotic, with systems crashes, lost communication between PDA and server and much more. It demonstrated that we needed to be far more careful about our management of the opening 20 minutes of the activity.

After a one-week interval – during which we put right as many of the glitches as we could identify – the team de-camped to Cornwall for a week of trials. The 2nd trial was altogether a smoother affair. The system held up throughout – and we only had a small number of pda freezes. We sailed through the challenge of box 1,2,3,4 and on each morning we completed the allotted sections of the activity. So by the end of the 2nd morning we had – for the 1st time – run the activity right through to the end. As a result, we were dealing with far more of the photography and sound files – 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.

It was clear by the end of the 2nd trial that the training time for students to learn the features of the ***e-scape*** application would have to be at least an hour and a half. With this starting point, the 3rd trial went very smoothly. We put in place all the protocols, and by now we were becoming more comfortable with the set-up and management of the system. Throughout the two mornings of the activity, very little data went missing, and the system worked well with all students' work live in the website.

Overall, the trials did what they were designed to do. They enabled us to tweak 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 make this important step forward.

2.5 the *e-scape* pilot (June/July 2006)

To run the national pilot we had acquired four sets of kit;

- four class sets of PDAs (24 machines per set - for 7 groups of 3 learners with 3 spare)
- four administration lap-tops running the local network
- four sets of handling collections
- four sets of modelling materials

This enabled us to run the ***e-scape*** activity in four sites simultaneously – provided that we also had four administrators. Twelve schools took part in the pilot, and since we could run four at a time, they were clustered as follows:

Round 1 (June 5th-9th)

Round 2 (June 19th-23rd)

Round 3 (July 10th-14th)

At each centre the lead teacher had been trained in the operation of the system in one of the two regional training sessions (Birmingham and Newcastle) that we had run in May 2006.

The pilot week in any school had the following broad framework. We undertook a training session with the learner group to familiarise them with the ***e-scape*** application and the PDA and specifically with the sketching, text entry, voice recording and camera functions that are important in the ***e-scape*** activity. Thereafter the activity itself ran for two consecutive mornings providing a total of approx 6 hours of activity. The activity was run by one of the research team. The lead teacher in the school was present throughout the activity to observe and assist as necessary.

i. setting up the system

On arrival at the school, the activity administrators' first task was to set up the laptop, wireless router and the PDAs in such a way that the PDAs each communicated with the laptop via the router. It is important to recognise that the preparation of the hardware involved a serious level of personalisation. The laptop was programmed with the class names and details of the up-coming centre, and each of the PDAs was allocated to one of the individual learners. So when the learner switched on the PDA for the 1st time it would immediately come live with the interface *'please confirm that you are Sam Walker of group 2'* (there were 7 groups of 3 learners in each class). Since the PDAs each had sticky labels with the same names, there was rarely any difficulty with this arrangement.

ii. running the activity

Once the system was up-and-running, the activity typically ran extremely smoothly. Our previous experience with *Assessing Design Innovation* and all the trialling of the task that we had undertaken in early 2006 ensured that the activity provided a rich and engaging experience for learners. They were typically enthusiastic about using the PDA as a design tool and they invariably responded to the 6 hours of activity with wholehearted commitment.

As the teachers became familiar with the workings of the system they too typically lost their nervousness and started to take a part in the management of the activity '"I'll do the next bit" and "Can I have a go at running it". It would have been too much to expect teachers to run it from the start, but by the end of the two mornings it was commonplace for them to be comfortable with it.

Following the activity, all the hardware was returned to the web-managers who stripped off all the learners' work and mounted it in the website. This could have been done remotely from the schools, but to save administrators time it was decided to back up the data set using a USB stick and return the kit. All the non-digital learner outcomes resulting from the activity (models, booklets, questionnaires) were transported to and stored in the TERU offices.

iii. the website

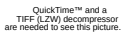
All the learner portfolios are in the website at the following address.

http://212.100.251.115/***e-scape***/

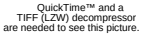
Because of the nature of the data on the site – with real names of schools and learners and including some photos and their voice files - access to the site is restricted to the research team, to DfES and QCA and to research staff at AQA and Edexcel.

Within the pilot we have a total of 12 schools for assessment and a total of 249 portfolios. Considering that the maximum possible number of portfolios was 252 (21 learners in each of 12 school), we consider that being only 3 short (for absence and other technical reasons) is a substantial achievement and we are immensely grateful for all the efforts of the schools, our administrator colleagues, and the hard and software support we received from the system developers.

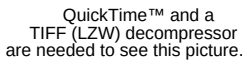
The website is organised through three layers of interface. Initially access is through the moderator log-in screen demanding personal access codes; then through a 'schools' page and finally to a class list that identifies names, UPN and group numbers etc.



Double clicking on the school will open the class list for that school, and double clicking on a learner name on that list will open their portfolio.



The portfolio is structured through the 24 sub-tasks of the 6 hour activity, with response modes of various kinds (drawing, writing, photographing and speaking) and with both individual and team-based purposes. Like the paper-portfolios that were the precursor to ***e-scape***, these web-screens provide a very real picture of the learners' evolving prototype and their thoughts about it along the way.



This snapshot of box 8 illustrates the richness of these data. The three photographs show the drawing up to that moment, and two photos of the model - from different angles. Clicking on the magnifying glass brings the images to full-screen size. The two sound files are the authentic recorded voice of the learner responding to two questions - what is working well? and what needs further development? – and together these provide a real insight into their understanding of their work.

It is important to note that this 'photo and sound file' routine recurs throughout the activity – essentially once an hour for the 6 hours. At least three significant things result from this. First, they get better – more articulate – in describing their work and the circumstances surrounding it. Second, the routine – taken together – leaves a real-time visual/audio evidence trail that is quite unique in the assessment of performance in design & technology. Third, learners' approach to the task is enriched as they are more motivated , braver (take more risks), and think more deeply about what they are doing.

Finally, the review comment (below the sound files) is a reflection by the learner made at the *very* end of the 6 hours of activity. Looking back over their whole work, we invite them to think about what they might have done differently if they had known then what they know now. Sometimes these meta-cognitive responses are descriptive – as in this case – and sometimes they are deeply analytic of their own performance.

We are not aware of any equivalent system of real-time, dynamic, e-portfolio assessment for any subject in any country. We believe this to be a world first. The 249 rich portfolios that inhabit the web-site have now become the focus of our work in the project. And the next major challenge was to operationalise the assessment process.

3 Making assessments

We are aware that the excitement generated in school trials about the learning potential of ***e-scape***, cannot be allowed to divert us from the assessment focus of this project. We have therefore been attending to this, and have explored a new approach to assessment that seems to offer much potential.

Through the Awarding Bodies - initially Edexcel – we have made contact with Alistair Pollitt, who was at one time the director of assessment research at the University of Cambridge Local Examinations Syndicate. He has drawn our attention to a system of assessment that we might call 'differentiated

pairs' since the assessment process involves making a single judgment about which of two scripts is the better. After a series of discussions with QCA, Edexcel, AQA, Alistair and the TERU team, we decided on a trial of the system using our existing archive of paper-based performance from the *'innovating assessment'* project.

Conventionally, assessment is undertaken by drawing up a list of criteria for the performance; allocating a block of marks to each criterion; and judging individual pieces of work to decide how many marks to award it for that criterion. This involves a judgement of the individual piece of work against the criteria, but interestingly, in the words of a recent book on the psychology of judgement, "There is no absolute judgment. All judgments are comparisons of one thing with another" (Laming, 2004). In other words, all judgements are relative. And in that case why not compare the work *directly*?

Based on this idea, an alternative approach emerged from the psychophysical research of Louis Thurstone, and specifically on his Law of Comparative Judgement (, 1927). The essential principle in this law is that, whenever a judge compares two portfolios (using their own personal 'standard' or internalised criteria) the judge's personal standard cancels out. The greater the true difference between the quality of the two portfolios the more likely it is that the better one will win each time they are compared. Thus a large set of comparisons does more than just generate a rank order; the relative frequency of success of one performance against another also indicates how far apart they are in quality.

> *Statistical analysis of a matrix of comparative judgements of 'scripts' can construct a measurement scale expressing the relative value of the performances. The result of comparisons of this kind is objective relative measurement, on a scale with a constant unit. Furthermore, if a few scripts that have already been agreed to represent grade boundaries – perhaps from a previous sitting of the examination – are included in the comparisons, the whole process of marking, grading and comparability of standards can be replaced by the collection and analysis of paired comparative judgements. (Pollitt 2006)*

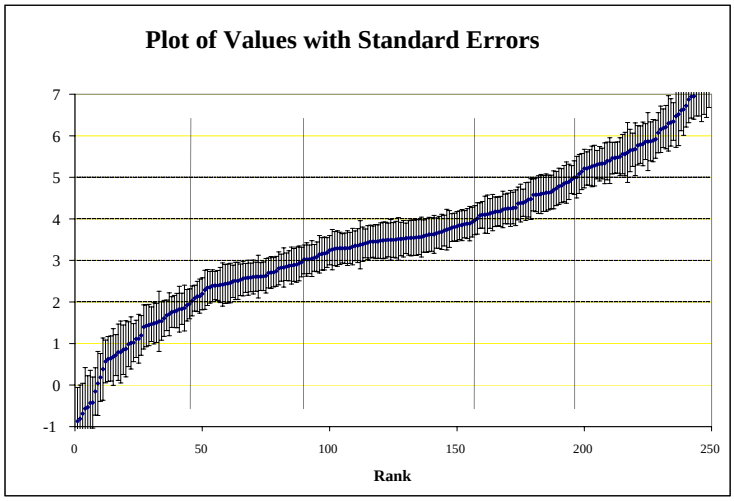
In a trial assessment – based on a set of work that we had formerly marked through conventional procedures - we showed that there is a strong relationship between the parameters derived from the comparative pairs judging and the marks previously awarded. The value of R^2 was 0.81, corresponding to a correlation of 0.90 between two linear variables, as high as could be expected in a case like this. Armed with such strong evidence, we felt confident to embark on the assessment of the whole ***e-scape*** sample using the Thurstone / Pollitt approach of comparative pairs judgement.

The judging process for the main body of ***e-scape*** data was designed in three rounds, using 7 judges. Round one involved each of the judges reviewing 140 pairs and deciding on the 'winner' in each case. The general response of judges was that the early pairs (say the 1st 20 or 30) took as long as 10 minutes per pair to decide, but gradually we got quicker. This speeding-up resulted in part from being more skilled in working our way around the web-based portfolio, and in part from the fact that the pairings inevitably threw up repeats. Having got properly inside a piece of work at the 1st time of asking, it took

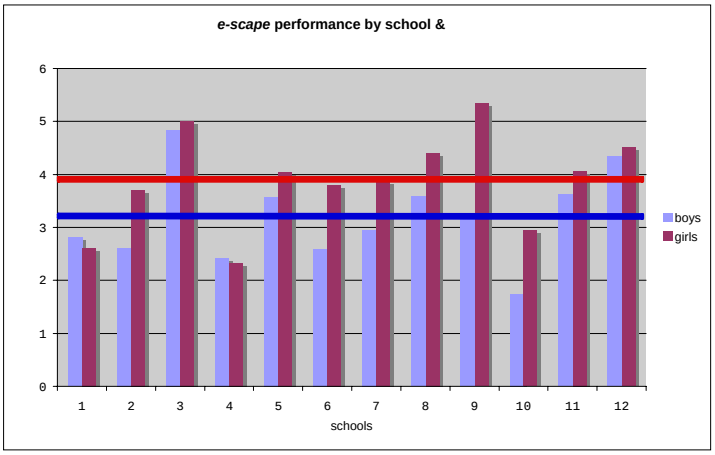
only a much briefer scan 2nd time around to remind us of its qualities. By the end of the 140 pairs we were typically doing each pair in 2 minutes.

Having completed round 1 judging, the results were analysed into a rank order and round 2 pairings were then selected to refine and confirm the order. The pairs this time tended to be closer together, comparing (for example) something like a B with a C, whereas round 1 (being random) had just as frequently required us to compare an A with an E. The relative difficulty of these round 2 decisions was offset by the familiarity (by now) with much of the work. Generally round 2 was quicker than round 1. Round 3 of the judging was then focused on enriching the data (so at to assure reliability) at the grade boundaries (A/B/C/D/E).

The complete set of results can best be seen in the chart here. In it, the portfolios have been sorted into order and are shown with their standard errors. In formal statistical terms 68% of the portfolios' "true" values will lie within one standard error of the reported value. Vertical lines are drawn through the five grade boundaries (A-E) to show how many pupils would fall into each grade. The analysis of the judgements also gives a traditional indication of the quality of the measurement process and 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.



When the scale is analysed by schools and gender the results are as shown here. Girls mean performance is (once again) higher than boys, and performance varies substantially across the 12 schools in the pilot. Some schools (esp 9) show a marked gender effect while in others (eg 3,4,12) it is almost eliminated.



4 Issues Arising

It is important to remember that there are two major innovations in project ***e-scape***. First we have created and tested a prototype system that allows students to operate with hand-held digital tools in the classroom and that enables their resulting portfolio to emerge dynamically in a secure website. This alone has massive repercussions for teaching, learning and assessment in schools. Beyond this however the second major innovation is that we have transformed the conventional view of assessment from one in which we allocate marks to an approach where we simply compare pairs of portfolios, judging in each case which is the better of the two. No marks are necessary for this direct comparison, but with enough comparisons it is possible to create a scale of performance that properly reflects the different levels of capability of the students. The same criteria of excellence can be applied – but we do not allocate (and then add up) numbers. A mass of issues arises as a result of the success of the prototype and of the subsequent 'pairs judging' of the web-based portfolios. In the space and time available, I have chosen to group them into three.

4.1 concerning classroom activity

All the teachers in the 12 schools of the pilot study were vociferous about one matter that was (they claimed) absolutely self-evident. Whilst ***e-scape*** is undeniably an *assessment* project, the most powerful consequences that they could see were about *teaching* and *learning*.

The idea of learners operating freely on hand-held computers – as a regular part of the learning environment – opens up all kinds of communication and enrichment opportunities. They could readily share their work with team-mates and teachers; they could operate at home and as well as at school; on field trips and at the theatre; collecting photographic scrap books/mood boards; recording interviews with users and experts; and all their work could be monitored dynamically as it emerged in the prepared web-space.

Moreover, this mobile technology is in the back pocket – with them always – not just when you go to the computer suite and sit at a 'terminal'. The design & technology environment is very challenging for normal computers that don't go well with dirt and dust and all the detritus of manufacturing. But a digital pocket-tool that can be got out to do a job and them returned to the pocket in the blink of an eye is a very different matter. And these tools are now very powerful. The PDA we used for ***e-scape*** had more processing power than the 1st generation of i-macs.

And the pedagogic consequences go way beyond design & technology. Science investigations, history explorations, language composition, geographic/landscape adventures can all be structured into activities that operate in an ***e-scape***-like manner.

4.2 concerning assessment

The innovation of Thurlstone/Pollitt style 'paired judgement' is made possible by the fact that all the portfolios exist live in a website. If it were not so then all the judging (of hundreds – and possibly thousands - of scripts) would be unmanageable. As it is, we just flick through the school and pupil

interfaces and call up *any* piece of work and compare it to any other piece. And our colleagues on the judging team can be calling up the very same pieces – but on their own machines at home. Access to the work is instant and ubiquitous.

And the question arises… 'who will be the judges?'. What if every teacher was a judge, comparing some of their own students' work with other pieces from other schools? Would teachers not benefit from this broader glimpse of students' performance? And what if students themselves were judges? Would they too not benefit from seeing other students' work and being able to compare and contrast strengths and weaknesses?

The natural tendency is to suggest that this leads (at best) to unreliability and (at worst) opens up the possibility of dishonesty, but the system deals with that, and in two ways. First there are multiple judges looking at every piece of work, and second each piece of work is compared to many other pieces. In the pilot, every piece was compared to at least 16 other pieces and these comparisons were judged by at least 5 of the 7 judges. The consensual / democratic decision always outweighs the possibly biased judgement of an individual. Moreover it is possible not only to identify 'rogue' judges – who are seriously outside the consensual mean of the judging group – but also to eliminate their judgements and re-calibrate the scale. The reliability statistics we achieved with ***e-scape*** are far in excess of normal GCSE (age 16) assessments simply because of these multiple-judge / multiple comparison factors.

4.3 concerning the technology

The problems with the technology that emerged through the trials were typically heavily outweighed in students' minds by the advantages. But we should not ignore the problems. Several students talked of the difficulty of drawing on a smallish screen; of the 'fiddly nature of the styles; and of not being able to see all the website 'at a glance' on the same screen. Various kinds of *scrolling-screen* were developed for the pilot, but all had weaknesses in relation to existing paper-based practice.

But the technology is evolving rapidly. "Projection" keyboards (the size of a box of matches) make it possible to project an infra-red, full-size, 'querty' keyboard and make it 'live' on an ordinary wooden desk surface. Equally, match-box size projectors linked to the pda make it possible to project the contents of the pda – and the linked website – onto a big screen or wall. So in a year or two (at most) the PDA will just provide the *processing and communication* facility. The input and output devices will be of our choosing, can be customised to the activity and auto-linked by blue-tooth.

5 Next steps

It is already abundantly clear that the prototype system we created for ***e-scape*** phase 2 has worked remarkably well despite the fact that the two major innovations are both in a rough prototype form.

The first element of innovation in ***e-scape*** (working with hand-held technologies linked to the website) was based on a digital 'lash-up' of hard and software systems that already existed. The PDA already had a draw / write / speech and photo facility, and the ***e-scape*** application wove them together into a digital package that triggered the right ones at the right times. Equally, the website system already existed but had to be reconfigured to accept the files coming from the PDAs. The ***e-scape*** application is essentially digital selotape; a prototype of a system that could work far better – much smoother – if it was based on specially created programming at both ends of the system (the device and the website).

Similarly, the second element of innovation (assessment using Thurlstone / Pollitt pairs) is currently used primarily for inter-board reliability and comparability studies. The system had to be seriously re-configured – and even somewhat re-conceptualised - to adapt to this new role, and Pollitt is honest in his analysis of the misjudgements he has made in that process. But it has worked, and we can see many ways to modify and streamline the process for this new role of what might be called 'front-line' assessment.

We have now launched phase 3 of project e-scape funded by Becta and running from Sept 1st 07 to March 31st 09. There are two central challenges for the new project:

• transferability … will it work for other subjects?

• scalability … can it be taken from a research project and scaled up to operate as a national system of assessment run by Awarding Bodies?

The prevailing mood of challenge to coursework assessment reinforced the importance of the work, since ***e-scape*** activities are effectively school-based, structured, time-limited projects that provide terrific evidence of process skills and render highly reliable assessments. Watch this space.

References

Department for Education and Employment (DfEE) 1999 Design and technology: The National Curriculum for England: Department for Education & Employment (DfEE) and the Qualifications and Curriculum Authority (QCA). London

Department for Education & Skills (DFES) 2003 Survey of Information and Communications Technology in Schools Oct 2003 HMSO - available at http://www.dfes.gov.uk/rsgateway/DB/SBU/b000421/index.shtml

Haste H. 2005 Joined-up texting: The role of mobile phones in young people's lives Haste H. Director of Research: Nestlé Social Research Programme Nestle - available at http://www.mori.com/polls/2004/nestlesrp3.shtml

Laming, D. (2004) Human Judgement: the eye of the beholder. London, Thomson.

Office for Standards in Education (OFSTED) 2004 ICT in schools – the impact of government initiatives: Secondary design and technology May 2004 OFSTED publications - available at http://www.ofsted.gov.uk/publications/index.cfm?fuseaction=pubs.summary&id=3649

Office for Standards in Education (OFSTED) 2004 (ii) ICT in schools: The impact of government initiatives 5 years on May 2004 OFSTED publications - available at http://www.ofsted.gov.uk/publications/index.cfm?fuseaction=pubs.summary&id=3652

Pollitt, A (2006). Paired-comparison scale creation for e-scape: a report from Alistair Pollitt to TERU

Thurstone, LL. (1927) A law of Comparative judgement. Psychological Review, 34, 273-286